

UC Berkeley

Dissertations, Department of Linguistics

Title

A Typological Evaluation of Celtic/Hamito-Semitic Syntactic Parallels

Permalink

<https://escholarship.org/uc/item/8p00g5sd>

Author

Gensler, Orin

Publication Date

1993

A Typological Evaluation of Celtic/Hamito-Semitic Syntactic Parallels

by

Orin David Gensler

B.A. (Harvard University) 1972

M.A. (University of California at Berkeley) 1980

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Linguistics

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor Charles J. Fillmore, Chair

Professor Gary Holland

Professor Johanna Nichols

1993

The dissertation of Orin David Gensler is approved:

Charles J. Fillmore May 17, 1993
Chair Date

Gary B. Holland May 18, 1993

Shanna Nichols May 17, 1993

University of California at Berkeley

1993

A Typological Evaluation of Celtic/Hamito-Semitic Syntactic Parallels

Copyright © 1993

by

Orin David Gensler

Abstract

A typological evaluation of Celtic/Hamito-Semitic syntactic parallels

by

Orin David Gensler

Doctor of Philosophy in Linguistics

University of California at Berkeley

Professor Charles J. Fillmore, Chair

For 90 years Celticists have known of a long list of features whereby Insular Celtic syntax deviates from the Indo-European norm, but agrees with Semitic, Egyptian, and/or Berber (hereafter "Mediterranean Hamito-Semitic"). Such features include VSO macrotype, conjugated prepositions, relative clause formation, subject-verb nonconcord, article placement in genitives ("house the-man"), verbal noun (VN: object in genitive) instead of infinitive, periphrastic tenses (BE+Prep+VN), initial mutations --- some 17 altogether. Struck by such resemblances, scholars proposed a variety of substratal or areal explanations: a prehistoric migration from North Africa to Britain (Morris-Jones, Pokorny), or a pre-Indo-European "Eurafrican" linguistic area (H. Wagner). These proposals, never well-received, have consistently been attacked as speculative and unprovable.

The crux is the question of coincidence. Is there any basis beyond hunch for judging whether the ensemble of Celtic/Hamito-Semitic (CHS)

similarities is likely to represent chance parallel development? And do the features show any natural typological affinity for one another?

This study opens up a new, empirical approach to the problem: a world-wide sampling of 64 languages to determine which non-CHS languages are most CHS-like (and how much so, and where), which CHS features are rare "exotics" globally (i.e., a minority population "bucking a trend" vis-a-vis a dominant majority), which show intercorrelation, etc.

The results are illuminating. Nothing remotely close to the CHS type recurs elsewhere. The "best matches" (and not very good ones) come from all over the world. But Africa (especially West Africa) is most CHS-like of all continents, arguing that the CHS type has a natural home in Africa; by contrast, the type is radically out of place in Europe. Nearly half the individual CHS features are exotics, making coincidence maximally unlikely as an explanation. Several features do show moderate intercorrelation, but not the exotics.

The CHS type is thus a highly unusual configuration globally, and far more at home in Africa than Europe. Independent parallel development can never categorically be disproven; but the coincidence would be a remarkable one indeed.

Charles J. Fillmore

To my mother

and

in memory of my father

Contents

Preface	vii
Chapter 1: Introduction and methodology	
1.1 Overview	1
1.2 Features and languages	5
1.3 An unpopular question	6
1.4 Evidence in historical linguistics	8
1.5 Contact: Pro and con	13
1.5.1 Two prehistoric scenarios	13
1.5.2 The case for contact	17
1.5.3 Extralinguistic evidence	19
1.5.4 The case against contact	26
1.5.5 Absence of lexical resemblances	30
1.6 The typological method	33
1.6.1 The concept of typology	34
1.6.2 Typological features and their values	38
1.6.3 Skewed distributions and typological inference	39
1.6.4 The survey	45
Chapter 2: The history of the problem	
2.1 Beginnings	47
2.2 John Morris-Jones	57
2.3 Julius Pokorny	69
2.3.1 Theoretical background	70
2.3.1.1 Substratal explanation	70
2.3.1.2 Theoretical problems in substratal analysis	78
2.3.1.3 Typology, innere Sprachform, and national mentality	85
2.3.1.4 Africa and Hamitic	91
2.3.2 Points of similarity	93
2.3.3 Reactions	112
2.4 Heinrich Wagner	118
2.4.1 Typology: terminology and constructs	122
2.4.2 Sprachgeographie	129
2.4.2.1 Wagner as linguistic geographer	129
2.4.2.2 Sprachgeographie and history	140
2.4.2.3 Celtic/Hamito-Semitic as a study in Sprachgeographie	145
2.4.3 Points of similarity	152
2.4.4 Reactions	175
2.5 Other recent work	184
2.5.1 G. B. Adams	184
2.5.2 Karl Horst Schmidt	186
2.5.3 Other studies	189
Chapter 3: The roster of features surveyed	
3.1 Methodological requirements	192
3.2 Features: Hamito-Semitic, Celtic, Indo-European	199

Chapter 4: Sources and sample	
4.1 Orientation	251
4.2 Grammar books and the problem of silence	252
4.3 Commensurability, reinterpretation, and hedged categories	255
4.4 The language sample	263
Chapter 5: Scoring and feature design	
5.1 General principles	270
5.1.1 Privative vs. equipollent features	271
5.1.2 Bivalency	278
5.1.3 Scoring as a matter of degree	284
5.2 Features	289
Chapter 6: Results and analysis	
6.1 The raw data	335
6.2 Whole-language evaluation: summing the features	340
6.3 Distribution of individual features	355
6.3.1 Distribution, skewing, and multivalency	355
6.3.2 Features	361
6.3.3 Skewed-exclusive distribution of the CHS type	373
6.4 Areal patternings	377
6.5 Diachrony	385
6.6 Feature independence and correlation	394
6.6.1 Correlations between features	395
6.6.2 Correlations involving types	414
6.6.3 Detailed assessment of selected correlations	418
6.6.3.1 Copying as an oblique relative-clause strategy	418
6.6.3.2 Verb-first word order and nonconcord with verb	420
6.6.3.3 Predicative particle and adpositional periphrastic	421
Chapter 7: Conclusions	
7.1 Results	425
7.2 Two perspectives on substratal methodology	427
7.2.1 Archaeological antidiffusionism: The megaliths	427
7.2.2 Thomason and Kaufman on substratal explanation	430
7.3 Criticisms and responses	435
7.4 Chronological anomalies in a substratal account	442
7.5 Assessment	456
General bibliography	464
Bibliography to Appendix 2	498
Appendix 1: Language sample	
Alphabetical list with sources	509
Sample by areas and genetic groupings	511
Map	513
Appendix 2: Data summaries for languages surveyed	514

Figures, Tables, and Map

Figure 1	36
Figures 2, 3, 4	41
Table 1a: Uncompressed raw scores	336
Table 1b: Compressed raw scores	338
Table 2a: Three whole-language scoring schemes (sorted Col 1)	344
Table 2b: Three whole-language scoring schemes (sorted Col 3)	346
Table 3ab: Absolute statistics (features)	356
Table 3cd: Percent statistics (features)	357
Table 3ef: Percent of non-N/I (features)	358
Table 4: Skewing within the features	374
Table 5a: Areal averages	379
Table 5b: Areal percentages (+)	380
Table 6: Diachronic data	386
Table 7a: Correlations of equipollent features	400
Table 7b: Correlations of privative and equipollent features	406
Table 7c: "Correlations" of privative features	411
Table 8: Clusterings of + and - scores	415
Map	513

Preface

The puzzle at the heart of this study has been on my mind more or less since the day when, as a Semitist with a growing interest in typology, I began studying Welsh and was struck by its uncanny similarity to Arabic and Hebrew. If my approach to the problem differs from my predecessors', it is a reflection of the difference in background. Notwithstanding a decade-long concern with Celtic, my main language area has always been Semitic; thus I approached the problem exactly backwards from those who had worked on it before me. Working as a typologist in the American tradition, my direct concern is not with substratal and areal linguistics but chiefly with descriptive crosslinguistic variation per se. The worth of substratal and areal explanation is patent when used carefully; it is equally obvious how easily such explanations can be misapplied.

Previous researchers on this problem sometimes convey the impression of having known the answer in advance. I did not. I undertook this study partly in the hope of finding out an answer to this particular concrete question, partly in an attempt to work out a legitimate methodology for arguing for prehistoric contact on the basis of structural similarities. Throughout this investigation, contact, typological affinity, and coincidence have all represented very real possibilities

of explanation for the Celtic/Hamito-Semitic cluster of resemblances, and I have bent over backwards to be both fair and critical to substratalists and their opponents alike. Indeed, the writing of Chapters 2 and 3 was all but complete before any of the actual typological investigation had begun.

I would like to extend my thanks to my dissertation committee, Charles Fillmore, Gary Holland, and Johanna Nichols; to Arwyn Watkins, John Koch, and Stefan Schumacher, who read and commented on parts of the first half of this dissertation, and to Eve Sweetser, who has been through much of it; to Joe Eska, for help on Continental Celtic; to Kathryn Klar and Annalee Rejhon, my Welsh teachers; and to the late Brendan O Hehir, my Irish teacher, who warmly (and skeptically) encouraged me in this study before his death. Thanks also to the many scholars with whom I studied Semitics in Berkeley and Jerusalem, and to my Egyptology teachers in Berkeley. Many friends shared with me their special knowledge of particular languages; they are mentioned in the relevant language summary in Appendix 2. Several scholars in Wales very kindly helped me try to figure out whether John Morris-Jones really did have a change of heart; others advised with bibliography on matters beyond my own ken.

Chapter 1: Introduction and methodology

1.1 Overview

This study has a narrow focus and a broad one. Narrowly conceived, it deals with two groups of languages. The first is Insular Celtic, comprising the Celtic languages of the British Isles (notably Welsh and Irish), including Breton but excluding the inscriptional remains of Continental Celtic. The second is a much broader grouping comprising Semitic, ancient Egyptian, and Berber --- three branches of the Afroasiatic superfamily to which I will collectively refer by the areal label "Mediterranean Hamito-Semitic", or "Hamito-Semitic" (HS) for short.¹ These two groups of languages, genetically unrelated at any recoverable time depth,² share a large number of striking syntactic commonalities --- features which, moreover, are fundamentally alien to Indo-European, the parent family of Celtic. The problem represented by this puzzling agreement and disagreement has been exercising Celticists for almost a

¹ In earlier work, especially among European scholars, "Hamito-Semitic" has been the name applied to the entire group now known (since Greenberg 1966b) as Afroasiatic, made up of Semitic, Egyptian, Berber, Chadic, and Cushitic (Omotic is often split off from Cushitic as a distinct branch). In this study, the term is intended strictly as a convenient areal label for a subset of that superfamily, and nothing more. In particular I intend no claim that Semitic, Egyptian, and Berber form a genetic subgroup. They certainly are reasonable candidates for areal subgrouping, both geographically and (as we will see) structurally.

² I ignore here the "Nostratic" hypothesis of a very deep genetic link between Semitic and Indo-European as a whole.

century. Convinced that such an impressive roster of similarities could not be accidental, scholars --- notably John Morris-Jones, Julius Pokorny, and Heinrich Wagner --- have proposed a variety of scenarios involving prehistoric contact between the groups in question. Right or wrong, these prehistoric reconstructions are hypothetical constructs, with little solid backing from history or archaeology. To an overwhelming degree they rest on purely linguistic arguments, and will stand or fall on the strength of those arguments. The majority of historical linguists, methodologically ill-disposed to entertain such substratal accounts, have rejected them as unprovable. There is near-unanimous consensus that the phenomenon per se is real and striking;³ the differences center on matters of interpretation.

The broader focus of this study takes in the entire world. For there are aspects to the Celtic/Hamito-Semitic (hereafter "CHS") problem which in principle cannot be resolved within the confines of Indo-European and Afroasiatic. When languages show similarities of any sort, there are four possible approaches to accounting for the resemblance. They can be ascribed to common genetic inheritance, or to contact phenomena in the broad sense of the word (including ad/sub/superstratal influence, areality, bilingualism, etc.), to a natural typological affinity among the similarity features, or finally to unmotivated coincidence. Within the context of the CHS problem, genetic inheritance is

³ There are dissenting voices, however: thus the following quote from Thomason and Kaufman's excellent book on language relatedness: "All the hypotheses that have been advanced about such a substratum [i.e. a Hamito-Semitic substratum underlying Insular Celtic] ... rest on such tenuous historical and linguistic evidence that the chances for a convincing proposal in this area seem remote" (1988:112). See sec. 7.2.2 for detailed discussion of Thomason and Kaufman.

ruled out. The possibility of typological affinity has, to my knowledge, barely been raised (Hewitt 1984-85, Jongeling 1988(?) ms), let alone articulated as an explanation. Earlier explanatory attempts have been framed in terms of contact, or some similar construct such as ancient areality (Wagner). Those who reject these explanations sometimes do so because they favor one of the alternatives (usually coincidence), but more often on methodological grounds, out of a reluctance to entertain questions not susceptible of proof. And indeed, within the familiar scope of Europe and North Africa, the question is more or less unresolvable. Issues like coincidence or natural affinity cannot be dealt with in strictly local terms, looking only at the concrete development of particular groups of peoples and languages. They require a broader focus, an understanding of what is likely and unlikely to occur and cooccur in human language as a whole.

Linguists of all persuasions seek insight into human language as a whole, but there are salient differences in the route chosen toward this goal. For some linguists, broad-focus linguistic insight is something that emerges from linguistic theory. A reliable theory of language, presumably, could tell us much about what is and is not common crosslinguistically. Such theories, however, are probably utopian, and at the very least unverifiable. The alternative is an empirical one: look worldwide at what actually occurs, and see what the actual distribution of the features in question is. This path, that of linguistic typology, will be the one to be followed in the present study.

This study, then, has a twofold focus, which will expand and shift as the exposition proceeds. After the overview provided in the present

chapter, we begin in Chapter 2 with a comprehensive review of previous work in the field --- some of it "typological", to be sure, but not in the global sense referred to here. Chapter 3 provides a detailed list of 17 features⁴ held in common by most of the older CHS languages but deviating from the Indo-European norm. Chapter 4 discusses the general strategy and problems involved in shifting to a global perspective, and lays out a language sample of 64 languages which will be surveyed in depth for their behavior with respect to the CHS phenomena. Chapter 5 presents the scoring system whereby a complex variety of "exotic" linguistic phenomena, much of it transcending anything seen in the familiar languages of Europe, is reduced to scores between +1 and -1, depending on how well the given feature in the given language agrees (+) or disagrees (-) with the CHS phenomenon. This will make it possible to quantify the CHS type. In Chapter 6, the individual scores are combined and analyzed in various ways to show how well each language as a whole counts as a match to the CHS type, how the CHS type itself patterns geographically, and how unusual each individual CHS feature is from a global perspective. Chapter 7, finally, pulls the threads together.

The result, in a nutshell, is that prehistoric contact (in the broad sense) does appear to be the most likely way of accounting for the resemblances. I do not believe this claim can be "proven"; I advance it only as the most reasonable way of interpreting the evidence.

⁴ Several of the features are subsequently split into distinct sub-features in Chapter 5, yielding a final total of 20.

1.2 Features and languages

To orient the reader, the 17 features to be presented in Chapter 3 will be stated here in preliminary capsule form. The feature numbering used here will be standard throughout this work, except in Chapter 2.

1. Conjugated prepositions: [Prep + Pronominal Obj] is a single word
2. Word order: VSO, N-Modifier, Prepositions
3. Relative clause linker: invariant particle, not relative pronoun
4. Relative clause technique (oblique): copying, not gapping, e.g.
 "the bed that [I slept in it]"
5. Special form of the verb peculiar to relative clauses
6. Polypersonal verb (subject and object both marked)
7. Infixing/suffixing alternation: Object marker is infixed
 to the verb if there is a preverb, suffixed otherwise
8. Definite article in genitive embeddings may occur only
 on the embedded noun: "house [the-man]" = the man's house
9. Nonconcord of verb with full-NP subject: verb can fail to
 agree with the subject, depending on word order
10. Verbal Noun (VN: object in genitive), not Infinitive (object
 in same case as with finite verb)
11. Predicative particle: in copular or nominal sentences, the
 predicate is marked with a particle homophonous to a "local"
 preposition: "He (is) in a farmer" = He is a farmer
12. Prepositional periphrastic: BE + Prep + VN, e.g.
 "He is at singing"
13. DO periphrastic: DO + VN, e.g. "He does singing"
14. Notional adverbial clause expressed as "and" + finite clause

15. Nonfinite forms usable instead of finite main-clause verb
16. Word-initial change, expressing a variety of syntactic functions
17. Idiomatic use of kin terms in genitive constructions, e.g.

"son of sending" = messenger; "son of (the) land" = wolf

All these features occur in at least one older language (or Berber) in both Celtic and Hamito-Semitic; some occur in all the older languages; all are rare or nonexistent elsewhere in older Indo-European. The emphasis on "older" is important: the phenomena at hand, in almost every case, go back as far as we can trace the development of the languages. In fact we will focus exclusively on old Celtic and Hamito-Semitic languages here, with the single and unavoidable exception of modern Berber.

1.3 An unpopular question

The list just given is, impressionistically, rather amazing. It is much longer, for example, than the list of ten or so trans-Balkan resemblant features commonly adduced in support of the Balkan Sprachbund (see e.g. Sandfeld 1930:163ff., or Schaller 1975:101-2), the group which sets the standard for the kind and quantity of evidence needed to argue successfully for interlanguage contact on the basis of structural resemblances. The long roster of CHS features poses a genuine problem, and the questions it raises should be attractive ones to historical linguists. Yet for the most part, previous work on this problem has been carried out on the margins of historical linguistics. Archaeologists, too, have methodological reasons for avoiding the issue. And in

the world of Celtic studies, the question until recent decades has been very nearly taboo. There are three quite distinct reasons for this, all intertwined with the histories of the respective disciplines.

Within the field of Celtic studies, the CHS problem has been the victim of pervasive backlash to an early pseudo-scholarly tradition genetically rooting the Celts --- language, people, and pedigree --- in the ancient Near East. This pre-scientific tradition flourished down to the eighteenth century, and even today enjoys some popular following (the British Israelites, for example, believe that English is descended from Hebrew). In soundly rejecting this quasi-mythological antiquarianism, modern Celtic scholarship has in effect made the entire subject anathema. Any proposal of prehistoric contact with the Mediterranean world has run the risk of being dismissed as not merely wrong but fantastical.

A similar backlash underlies the attitude held by much of modern archaeology vis-a-vis problems analogous to the CHS puzzle. For decades, the dominant explanatory paradigm in European archaeology was diffusionism. Artifact types or styles found in non-Mediterranean Europe were explained by invoking putative typological parallels in the Aegean or the Near East, which were then supposed to have diffused to the less "civilized" parts of Europe. With the coming of the radiocarbon revolution, this paradigm was revealed as factually untenable, and a reaction set in against any form of diffusionist explanation (see e.g. Renfrew 1979:15-19). The CHS problem, however, requires one to take the possibility of diffusion --- what I have called "contact" --- very seriously, and to look closely at a whole set of typological parallels, not

archaeological but linguistic. This runs against the dominant current in modern archaeology. (See sec. 7.2.1 for a detailed critique.)

Within the world of Indo-European studies, the issues are more substantive. The special status of the comparative method, as a uniquely reliable tool for valid inference on linguistic prehistory, has tended to prejudice many scholars against historical problems which do not come under its purview. We will examine this point in much greater detail in the next section. Suffice it to say here that there is a potent methodological mistrust of historical conclusions based purely on structural commonalities across languages. There is nothing inevitable about this; historical inferences based solely on linguistic data of other sorts are accepted confidently. Almost no one would think to demand extralinguistic corroboration of historical conclusions based on the comparative method; to the contrary, it is accepted that comparative linguistic results may serve as the basis for reconstructing the protohistory of peoples, as in the Indo-European proto-homeland question.⁵ Of course, with no method in hand for evaluating the evidential value of structural resemblances, such radical skepticism is quite proper. A major goal of this study is to demonstrate such a method in operation.

1.4 Evidence in historical linguistics

One of the paramount themes presiding over the development of historical linguistics has been comparison of languages --- and, in

⁵ One need only think of Navajo vis-a-vis the Canadian Athapaskan languages, or Wiyot and Yurok vis-a-vis Algonquian, or Malagasy vis-a-vis the rest of Austronesian.

particular, the effort to determine how to evaluate crosslinguistic similarities, to appreciate what their diagnostic status as evidence might be. As remarked, linguistics has established various possible causes for observed similarities: they may be due to genetic relatedness, to contact and borrowing, or to "universal" considerations such as onomatopoeia and typological affinity --- or, finally, they may be written off as coincidence. Of these various types, comparative/historical linguistics has always focused overwhelmingly on the first, namely genetic relatedness, especially taken in opposition to borrowing. A powerful methodology was developed, the comparative method, which provided a uniquely trustworthy means for attributing a particular point of crosslinguistic resemblance to common genetic inheritance. The great reliability of this method stems largely from a built-in control factor: an emphasis on quirky similarities, and to that end a focus on resemblances involving both sound and meaning, i.e. involving actual morphemes.

The prime "competitor" of genetic relatedness has always been borrowing, and the neogrammarians and their successors were at pains to keep the two apart. The establishment of sound correspondences and an insistence on their regularity of operation were one way to help weed out resemblances due to borrowing, inasmuch as borrowed words (except for extremely early loans) typically can be expected to violate sound laws. A focus on "basic vocabulary" was another important methodological tool to the same end, as was also an emphasis on correspondences involving not just roots but whole words, including derivational and/or inflectional material in both languages.

The main thrust of the comparative linguistic enterprise, in both theory and methodology, has thus largely been aimed at using sound-meaning resemblances to establish and articulate genetic inheritance, and to discriminate between it and borrowing. Much less attention tends to be directed at the validity or reliability of the resemblances *qua* resemblances. It is not typically the case that the historical linguist, when dealing with a pair of resemblant word forms in two languages, must make it his or her first task to defend the resemblance against charges of coincidence. Comparatists have of course been aware of the problem of coincidence, and notorious examples of accidental lexical look-alikes (Sp. *mucho*, Eng. *much*; Pers. *bad*, Eng. *bad*) have come up for discussion (for example, Meillet 1964:380) --- but usually only in passing. An article such as Bender's "Chance CVC correspondences in unrelated languages" (1969), devoted expressly to the problem of coincidence in lexical similarities, is a rarity (see also Oswalt 1991, in a slightly different vein).

What is the evidential value of resemblances that do not involve both sound and meaning --- interlanguage resemblances in structure and system? Historical linguistics has never known quite what to make of such similarities. In genetic linguistics, varying views have been held on the importance of structural resemblances for language relatedness. A linguist such as Franz Boas believed strongly that "inner form" was conservative and would reveal much more about language history than other determinants.⁶ Few would hold to such a view today, for the study

⁶ "Lexicographic borrowing may proceed to such an extent, that the substance of a language may be materially changed. As long, however, as the inner form remains unchanged, our judgment [on the 'genealogical question'] is determined, not by the provenience of the vocabulary, but

of areal linguistics has shown that structural features are subject to diffusion just as words are. Far from having a privileged position, in fact, purely structural resemblances ought to be a much weaker diagnostic of genetic relatedness than lexical resemblances: they are not subject to the tight twofold control of a similarity in both sound and meaning, and they cannot be regimented in sound laws. There has, of course, been considerable work in reconstructing (aspects of) the proto-syntax of languages, particularly Proto-Indo-European --- but this presupposes genetic relatedness, and does not contribute to our understanding of the diagnostic value of purely structural resemblances as evidence for historical inference. Very few modern studies have explored the use of pure syntactic or morphosyntactic patterning, divorced from sound-meaning correspondences, as an argument for genetic relatedness: two salient examples are Fokos-Fuchs 1962 for Ural-Altaic, and Hymes 1956 for Na-Dené.⁷

The above discussion has identified two salient trends in historical linguistics as regards its treatment of crosslinguistic resemblances:

1) A focus on genetic relatedness, especially in opposition to borrowing, with a concomitant deemphasis on universal factors or coincidence.

2) A bias in favor of resemblances involving both sound and meaning, with a concomitant deemphasis on purely structural resemblances. This

by that of the form" (1917:3). Boas is drawing here on the long European tradition going back to Wilhelm von Humboldt; see sec. 2.3.1.3 below.

⁷ Hymes looks not at syntax per se, but at the syntagmatics of positional slots within the word; Athapaskan, Tlingit, and Haida agree closely in the order and overall meaning of positional slots, regardless of the phonological form of the morphemes filling those slots.

second trend is natural and justifiable given the first.

Yet the most basic conceptual split is not between genetic inheritance and borrowing at all. Both of these, after all, share a presupposition: that the peoples speaking the languages at some stage had a shared history, which we invoke in explanation of the resemblance. Logically, the fundamental dichotomy is rather between historical and ahistorical accounts of resemblances: can the given resemblance be taken as the result of "contact-or-relatedness" --- conceived of as a single historical macroconcept --- on the part of the two languages in question? Quite naturally, standard diachronic linguistics has devoted very little theoretical or methodological attention to this issue, for the form/meaning resemblance per se is typically so blatant (in the light of known sound laws) that there seems no need even to pose the question. In consequence, there is little in the way of established procedure or theory to fall back on when this question does come to the fore --- as in the present study.

There is, of course, one historical field where structural resemblances have long been acknowledged to have great value as evidence: areal or Sprachbund linguistics. For most linguists, indeed, the term "language area" definitionally involves the presence of shared structural features. But even here structural resemblances are only one factor contributing to the establishment of a linguistic area. Of overwhelming importance is the sheer fact of geographical proximity, and usually the co-presence of borrowed words.

The Celtic/Hamito-Semitic problem, accordingly, falls outside all the main currents of historical linguistics as outlined above. It does

not rely, in making historical inferences, on resemblances involving both sound and meaning, nor is it concerned with discriminating between genetic relatedness and borrowing; hence it cannot, on two counts, appeal to the comparative method. It does not involve geographical proximity; hence it cannot be dealt with in terms of the theory and praxis of areal linguistics. Yet the body of facts put forward as evidence is quite massive. What is unclear is its value as evidence, and what the evidence may point to. Does the resemblance entitle us to infer contact-or-relatedness on the part of Insular Celtic and Mediterranean Hamito-Semitic? Orthodox historical linguistics, as noted, has had little to say to this problem. Indeed, there never has been any method or procedure, nothing analogous to the comparative method, for establishing the diagnostic value of purely structural resemblances as evidence in historical argumentation.

1.5 Contact: Pro and con

1.5.1 Two prehistoric scenarios

In concrete terms, what does it mean to suggest that Celtic and Hamito-Semitic have a shared history? Two quite different scenarios have been advanced by scholars. The first, associated with Julius Pokorny but also advocated by Morris-Jones and his predecessors, is straightforward. On this view, a wave of people speaking a Berber-like language swept up from North Africa along the western fringes of Europe to the British Isles. When the Celts subsequently migrated from continental Europe to the British Isles around the 6th century BC, they found this Berberoid language awaiting them as a linguistic substratum. The Hamito-Semites were then absorbed into the dominant Celtic

population, but their language left an indelible structural imprint on the development of Insular Celtic, engendering fundamental syntactic changes in the language vis-a-vis the inherited Indo-European type. This is the substratal theory; it will be examined in detail in the discussion of Pokorny's work in the next chapter (sec 2.3.1.1, 2.3.1.2).

The second view, quite different from the first, emerges between the lines (though seldom explicitly) in the work of Heinrich Wagner. Wagner dubs his approach "sprachgeographisch", or areal. On this interpretation, languages of a type very different from that represented by Indo-European would have been spoken by the pre-Indo-European peoples of Europe and North Africa --- a linguistic stratum Wagner calls "Eurafrican". When the Indo-Europeans came into Europe from the East, this ancient linguistic stratum was overlaid almost everywhere by the newcomers' language. But its structural features survived on the margins, and left their imprint on the historically attested languages: in North Africa (Berber), in Britain and Ireland (Insular Celtic), and allegedly in Iberia (Basque). Here the contact between Celtic and Hamito-Semitic would have been ancient and tenuous, a loose linguistic interlinkage that would have put a common stamp on languages spread over a linguistic area of vast size. Wagner's focus, in fact, is seldom on the historical articulation of this scenario, but overwhelmingly on the synchronic description of the area per se. We will examine this areal theory in detail in sec. 2.4.2.

There is, of course, no guarantee that these scenarios are not merely exercises in creative fiction. However, both do embody a minimal presupposition about the genesis of the Insular Celtic linguistic type

which surely is right, and which I will take for granted throughout this study. That is the assumption that there was a pre-Celtic substratal language of some sort in Britain, and that the divergences between Insular Celtic and Indo-European are indeed due in considerable measure to this substratum,⁸ which was similar in type to what we know as old Insular Celtic. That some such language existed can be taken for granted; we know (or assume) on archaeological grounds that the Celts were not the earliest inhabitants of the British Isles, and indeed Pictish (or what little we know of it) gives indications of not having been Indo-European (Jackson 1955:155ff.). Given this assumption, the comparison in principle becomes not one between Hamito-Semitic and Insular Celtic, but between Hamito-Semitic and the posited pre-Celtic substratum as glimpsed through Celtic. And the question now becomes whether or not it is plausible to posit some historical connection between this substratal language and Hamito-Semitic on the basis of the observed structural similarities. The simplest form of such a connection would be one of identity (Pokorny's scenario); the more complex form would involve some kind of areal linkage (Wagner's view).

That there was in fact a migration of Celtic-speaking peoples from the mainland of Europe to the British Isles, where they overlaid some earlier population speaking a different language, would seem absolutely unproblematical: Celtic is an Indo-European language, and every one of the competing views on the Indo-European homeland question (see e.g. Mallory 1989, Renfrew 1987) places the IE homeland far to the east of the British Isles. Yet the issue is snarled in a web of puzzles.

⁸ So too Schmidt (1990:196).

When did the Celts come to Britain and Ireland? Did they come as a body, or in waves? Did the split into Brythonic and Goidelic Celtic evolve on the British Isles, or does the division have antecedents going back to the time of Continental Celtic? Just what is the linguistic relation between the Celtic dialects of the British Isles (Brythonic, Goidelic) and the Celtic dialects of the Continent (chiefly Gaulish and Celtiberian)? Does "Insular Celtic" represent a unitary genetic grouping? (Should we retain the notion of "unitary genetic grouping"?) And, most basic of all, did it really happen that a people who may be called Celts migrated to Britain, or was there rather a process of "cumulative Celticity" which gradually transformed Britain through slow diffusion from Europe? As Renfrew puts it, one might perhaps think of "England and continental Europe as starting on a more equal footing, and developing together that cumulative mutual Celticity which results in the position which we see at the time of Christ" (1987:246); the Indo-European penetration into Britain and Ireland would thus have occurred long before anything like "Celtic" could be said to exist anywhere, and "Celticity" (linguistic and cultural) would have arisen gradually and mutually on both sides of the Channel. All these topics have been argued hotly in the last few decades, and transform an apparently simple assumption into a complex problem.⁹

I have raised these issues largely to ignore them. I will uncritically accept that a group or groups identifiable as Celts and speaking a

⁹ For recent statements on "the coming of the Celts", see e.g. Greene 1966a, 1983, D. Ellis Evans 1986, 1991, Koch 1991; for the archaeological perspective, Mallory 1989:95-107, Renfrew 1987 (Chapter 9), Piggott 1983. The literature on the subject is enormous.

Celtic language(s) did move from Europe to the British Isles; Renfrew's view does not seem linguistically tenable to me. I will likewise accept the "terminus post quem non" used by most scholars, and say that the Celts were in Britain and Ireland at the latest by perhaps 600 BC (Greene 1983 lays out the position clearly). Within the context of this study, after all, it is not very important just when the Celts came over, or what the exact scenario was. Whether abruptly or gradually, as invaders or traders or settlers, what counts is that they did come over and that some other people was there before them.

1.5.2 The case for contact

In assessing the likelihood of prehistoric contact, it is instructive to compare the Celtic/Hamito-Semitic situation with that faced by an areal linguist (in the usual sense of the term, not Wagner's). If two or more neighboring languages show a shared structural feature, the unmarked assumption is that the resemblance is an areal feature stemming from a shared history. Linguists characteristically feel little compulsion to demonstrate or even argue that the resemblance is noncoincidental; typically the issue does not even arise.¹⁰ The reverse, surely, holds for two languages located halfway around the world from each other: very few linguists would think to ascribe a resemblance to anything but coincidence or typological universals.¹¹ But Celtic and

¹⁰ An exception is Hook 1983, with regard to areal features of India.

¹¹ Some linguists have boldly (or overboldly) attempted to infer a shared history from unsystematic lexical resemblances among languages at great geographical remove from each other --- Joseph Greenberg's recent work on Amerindian (1987), for example, or Otto Sadovszky's attempts to establish a link between California Wintun and Siberian Ob-Ugrian. Such

Hamito-Semitic occupy a peculiar halfway position --- geographically they are neither particularly close nor impossibly far apart. Thus the case for contact-or-relatedness (vs. coincidence or universals) is neither self-evident nor out of the question, and Pokorny's or Wagner's scenarios are neither obvious nor absurd. Whether one accepts them will depend on the strength of the arguments adduced.

Two considerations lend a strong degree of prima facie plausibility to the case for prehistoric contact. The first is that most of the characteristic Hamito-Semitic-like oddities of Celtic syntax seem to be found only in Insular Celtic, not in the much earlier Continental Celtic spoken widely in Europe during classical antiquity.¹² Indeed, this is clearer now than in Pokorny's day, for our understanding of Continental Celtic (in its various dialects) is far better than was his. The corpus has grown spectacularly since 1970, with major discoveries of long inscriptions from Chamalières, Larzac, and Botorrita.¹³ Of course, even this explosion of knowledge cannot change the fact that Continental Celtic is an inscriptional language, attested only fragmentarily; what we know of the language comes from glimpses, and much of the material remains enigmatic. Accordingly, we will not be able to treat every or even most features of Celtic/Hamito-Semitic resemblance from the

linguists make the characteristic methodological mistake of trying to argue directly for genetic relatedness, without thinking to first attempt the more modest argument for contact-or-relatedness vs. coincidence.

¹² Several of the features, to be sure, exist in nascent form on the continent; see feature-by-feature discussion in Chapter 3, and especially sec. 7.4.

¹³ See e.g. Koch 1983 and Schmidt 1976 for studies of Gaulish and Celtiberian syntax, respectively, and more recently Eska 1993 (ms).

perspective of Continental Celtic. The general point, however, appears valid: the overall syntactic cut of Continental Celtic differs only in minor ways from that of the other ancient inscriptional languages of Europe. The typological shift in Celtic syntax, accordingly, appears to be specifically linked with the move to the British Isles.

The second point concerns the nature of the resemblances themselves. In an argument for contact-or-relatedness based on similarities of any kind, whether structural or lexical, it is evident that resemblances have the highest value as evidence when occurring not singly but as part of an ensemble of other resemblances. For lexical similarities, little can be concluded on the basis of a few scattered words; the case for contact-or-relatedness becomes much better if many words are involved; better still, if the words all share a sound correspondence. For structural similarities, a single point of resemblance will seldom be informative; much better if the languages show many such similarities. Thus it is highly significant that Celtic and Hamito-Semitic share a very large number of similarities, and especially that many (not all) of these appear to be more or less independent variables. We will return to this issue in Chapter 6 (sec. 6.6).

1.5.3 Extralinguistic evidence

Obviously, a substratal account would be greatly strengthened if there were reliable nonlinguistic evidence to back up the linguistic similarities. Supporters have mustered as much information of this sort as they can; critics may argue for its relative paucity and thinness. The evidence appealed to comes from various fields: culture, religion,

history, archaeology, anthropology, and (perhaps most significantly) blood typing. In this section I will summarize, very briefly, what has been said on the subject. It should be emphasized that evidence of this sort, though tantalizing --- especially the blood-type data --- plays little direct role in the present study, concerned as it is purely with the evidential value of linguistic data. With no special competence in any of these fields, I am completely dependent here on secondary literature and the opinions of others, which I will simply report with little in the way of critical evaluation.

Proponents of cultural links between the British Isles and North Africa stress that one need not view early Ireland simply as an isolated appendage of Britain, nor Britain just as a corner of Europe, each island facing inward toward its larger neighboring land mass. One might instead look to the sea, and view Ireland as a kind of "traffic island, centre of a vast trade in boats up and down the Atlantic coasts from the Baltic Sea to the Straits of Gibraltar" (Quinn 1986:24).¹⁴ Quinn documents instances of cultural contact between the British Isles and North Africa going back to prehistory, ranging from the Barbary ape skeleton found in Ireland at Eamhain Macha (~200 BC),¹⁵ to a gold dinar inscribed in bad Arabic with the name "Offa" (an 8th-century Anglo-Saxon king) stamped on it, to a 9th-century Arabic-inscribed cross in southern Ireland, to 17th-century raids on Ireland by the Barbary pirates.¹⁶ Seen

¹⁴ Quinn's popular book, though not a work of "scholarship" in the orthodox mold, is the best overview of extralinguistic connections between the British Isles and North Africa that I have come across. It includes a bibliography of other relevant books and studies.

¹⁵ My thanks to James Mallory (p.c.) for this date, which has an error margin of perhaps a century either way.

¹⁶ Ape: 34; dinar and cross: 107-8; piracy: 38-56. On North African

in this light, the many cultural similarities with the Mediterranean need not be taken as historical freaks, the chance residue of penetration across the full breadth of mainland Europe, but as natural consequences of the islands' intrinsically maritime orientation. Quinn's book abounds in such points of resemblance: in the forms assumed by early Christianity in Ireland and in the eastern and southern Mediterranean, similar to one another but quite different from the mainstream Roman church (notably as regards doctrine and monasticism); in Arab and Irish music, especially the traditional Irish sean-nós genre; in art styles and motifs, especially early Christian iconography in Ireland and Egypt; and more.¹⁷ Quinn is hardly the first to have noticed such things; to take only one example, Pokorny (citing work by Wolfgang Philipp) had earlier alluded to "very strange parallels between Libyan and Insular Celtic customs, particularly referring to the frequent traces of matriarchate and the queer sexual morals to be detected in the Irish saga-literature, the position of the bards, etc." (1960a:233).

For the most part, the concrete details of cultural similarities and historical contacts between the British Isles and North Africa can provide little of direct relevance to the problem of Celtic/Hamito-Semitic linguistic connections. The chronology is wrong. Putative links with Arabs and Copts are centuries too late to be involved in any pre-Celtic substratum; they postdate, not predate, the coming of the Celts to Britain. Trade and cultural contacts in historical times would

pottery in Dark Age Britain and Ireland, see Thomas 1976; on direct literary contact between Ireland and Spain in the 7th century, see Hillgarth 1984.

¹⁷ Christianity: 115-16, 141-52, 172-73; music: 18-22, 28-29, 147-48, also Feehan 1981; art: 31-32, 78-79, 140-41, 153-56, 162-63.

at best have imposed a thin superstratum on preexisting Insular Celtic. Rather, the implicit point that emerges here is a methodological one. On Quinn's account, a sea highway linking Britain and North Africa emerges as far back as the historical record can take us. There seems no good reason why such sea contacts, if taken as a reality in historical times, should not also have existed in prehistory.

Classical sources do include one concrete indication of contact between the British Isles and ancient Mediterranean Africa: the well-known notion that the Phoenicians sailed as far as Britain in their search for tin. The source for this idea is Rufus Festus Avienus's geographical poem Ora maritima (English translation: Murphy 1977), a 4th-century AD Latin reworking of much older Greek sources, few extant, going back to an Iberian periplus (coastal sailing book). Avienus (lines 95-119) reports Carthaginian contacts, notably by the navigator Himilco (5th century BC?), with the "Oestrymnides", islands rich in tin and lead; from here it is two days' sail to the "holy island", where the "Hierni" live; the island of the "Albiones" lies nearby. Most scholars have taken these names as referring (respectively) to Brittany, Ireland (Hibernia), and Britain (e.g. Moscati 1968:181, Harden 1963:170-71). And there is no question that Britain was known to the classical world as a source of tin (Penhallurick 1986:139ff.). But it has been vigorously disputed --- as argued in Hawkes (1975:19-26), followed by Penhallurick (1986:123ff.) --- that Avienus's references necessarily point to Britain. Avienus's poem, Hawkes suggests, represents a jumble of several distorted traditions; the distances and times given in it are incompatible with a British interpretation; and the place names can instead be taken as referring, at least in the original periplus, to

locations on the Atlantic coast of Iberia, also sources of tin. Significantly, the British archaeological record includes Greek material but no clear evidence of a Phoenician presence (Harden 1963:171, Penhallurick 1986:129). As far as I know, the issue remains unresolved.

Archaeological evidence bearing directly on the question of prehistoric contact between the British Isles and North Africa does not appear plentiful.¹⁸ The prehistoric Barbary ape skeleton has already been mentioned, presumably a prestige gift from one ruler to another. Much later, there are indications of North African pottery of the late Roman period (400-800 AD) appearing in Britain and Ireland (Thomas 1976). If, however, we broaden the focus and ask instead about prehistoric sea links between Britain and the rest of western Europe, evidence of the most dramatic kind would indeed appear to exist: the megalithic tombs of prehistoric Europe, found not only in Britain and Ireland but recurring along the Atlantic coast from Denmark to Spain, along the northern Mediterranean coast, and in mainland France as well (see maps in Daniel 1963:26, Renfrew 1979:125). The tombs known as "passage graves", in particular, show a strongly maritime distribution: none are more than 30 miles from the sea (1963:97). In the British Isles, "the emphasis distributionally is on the western seaways" (1963:112).

Until a few decades ago, the Atlantic megaliths were accepted near-universally as a parade example of contact with the Mediterranean, and more particularly with the Aegean. Daniel sees the megalith

¹⁸ As Pokorny himself admitted (1960a:229): "Archaeology alone is quite insufficient to give us a proper idea of the earliest history of a country, particularly in Ireland, where there remains so much to be done in excavation work."

builders as traders, colonists, and metal prospectors, promulgating a burial type which ultimately had its origins in the eastern Mediterranean (136-37); he finds "good archaeological evidence of contacts between the Mycenaean world and the Early Bronze Age people of southern Britain" (130-31). But with the advent of tree-ring calibrated radiocarbon dating, and the "long chronology" it implies for early European prehistory, this view became factually untenable: the Atlantic megaliths, far from responding to Mediterranean influence, were in fact shown to predate them by a millennium and more (Renfrew 1979:passim, esp. 89-90 and 120-46). Renfrew argues instead for multiple local European origins, warning against an overreadiness on the part of "diffusionists" to appeal uncritically to typological resemblances as evidence for contact (1979:2).¹⁹

The Mediterranean coast of Europe is not yet Africa. But the distances are not great, and major megaliths exist on Malta, less than 200 miles from Tunisia (1963:85-89). As with the Atlantic structures, Renfrew argues here for local origin (1979:147-66). There are megaliths in North Africa too; Daniel passes over them in a few words, judging them to be a millennium later and ascribing them to independent parallel development (22, 81-82, 133). I do not know how these monuments have fared chronologically with regard to radiocarbon dating. As regards typological parallels, Quinn makes a point of noting, on North African stones and steles, the same sort of serpentine and concentric-circle

¹⁹ See Renfrew 1979 for detailed discussion of calibrated radiocarbon dating; and Renfrew 1979, 1983 (ed.) for the megaliths. MacKie 1977 presents a revised diffusionist account; see especially pp. 92, 161-62 for discussion of connections between Britain and the Mediterranean-North African world.

motifs characteristic of the megaliths of western Europe; he describes a circle of standing stones near Tangier which is "unmistakably a first cousin of Newgrange and Gavv'inis" (the famous megalithic tombs of Ireland and Brittany) (1986:71-74).

Early students of the Celtic/Hamito-Semitic problem tended to appeal to similarities in physical anthropology --- such features as darker skin color, short stature, and elongated skull shapes reminiscent of Mediterranean peoples.²⁰ More telling, however, is the dramatic resemblance in blood type between the inhabitants of the more remote western regions of the British Isles and of the central Sahara.²¹ In Western Europe, gene percentages for the A and O blood groups are roughly:

A=30, O=65

In certain western regions of Britain and Ireland, and also Iceland, the gene frequencies are roughly:

A=15-20, O=75-80

These percentages of A and O are, respectively, the lowest and highest in Europe. The two closest regions having comparably low/high figures are the Sahara desert --- Berber country --- and the Caucasus, specifically the small area at the eastern tip of the Black Sea.²²

²⁰ Thus e.g. Pokorny 1960a:229-34; for Ireland, see Lundman 1957; for Wales, Mourant and Watkin 1952:20-24; also MacKie 1977:91-92, 162.

²¹ For a linguistically oriented survey of the problem, see Holmer et al. 1961 (with maps); for Ireland, again Lundman 1957; for Wales, Mourant and Watkin 1952; for a global perspective, the studies of Mourant et al. (1958:265ff., 1976:maps [at end]). My summary draws on all these sources, particularly the maps. Note that the distribution presented on the 1976 maps is somewhat less sharply profiled than that suggested in the earlier works.

²² The Saharan regions in question stretch westward all the way to Senegal, where the language Wolof is spoken, one of the best linguistic

A final point. Holmer et al. note a general distributional fact about the O gene, and propose a plausible prehistoric reconstruction: "It seems that populations with a high O frequency tend to have a peripheral distribution reached presumably under the pressure of other populations with other blood group gene frequencies, ... [who] have forced the old populations with high O frequencies to move to different peripheral areas" (Holmer et al. 1961:18-19). This has echoes to the kind of Sprachgeographie proposed by Heinrich Wagner, alluded to above and discussed in detail in the next chapter.

1.5.4 The case against contact

The Celtic/Hamito-Semitic problem, as mentioned, has never enjoyed much popularity; that is even more true for the specific proposals put forward by Pokorny and Wagner. Neither Pokorny's substratalism nor Wagner's Sprachgeographie has been accepted by the mainstream of comparative linguists; most reactions have varied from agnostic or indifferent to actively derisive.²³ Such critics can point out that the "remarkable" similarity boils down to a purely subjective impression; surely, they feel, it could all be just a coincidence. In all

matches to the Celtic/Hamito-Semitic type outside Celtic/Hamito-Semitic itself. The Caucasian area is the home of the Northwest Caucasian language Abkhaz, another good linguistic match to the Celtic/Hamito-Semitic type. (See Chapter 6 for discussion of these languages.) Note in passing that Basque has the lowest percentage of the B gene in Europe; Celtic and Hamito-Semitic show average values for B, except for very high values in remote mountainous regions in Wales.

²³ For the latter view, see Calvert Watkins's reference to "a variety of hypotheses of the presence of sundry ill-defined non-Indo-European substrata in Celtic, particularly on the shadowy elfin-populated Emerald Isle" (1962:1).

likelihood, too, equally strong clusters of resemblances could be found between Insular Celtic and other languages. And the quality of the evidence plainly cannot match that employed in standard genetic argumentation and demanded by the stringencies of the comparative method. Of course, the advocate of contact has just as powerful a conviction of the overwhelming cumulative weight of the complex of structural resemblances, of their validity as evidence (witness Pokorny's repeated dictum that "es kaum ein Zufall sein wird"); but he cannot go beyond intuition either. The significant issue here --- more than the matter of who is right --- is that the question is being prejudged by both sides, and on no principled basis.

One of the telling arguments offered by critics of the substratal view is that substratal explanation is unnecessary. Numerous Celticists have devoted detailed attention to demonstrating how the undoubted structural oddities of Celtic vis-a-vis "normal" Indo-European can actually be seen, individually, as the continuation of possibilities already present in Indo-European. Calvert Watkins, among others, argues that much of the anomalousness of the Celtic verb --- its sentence-initial position, its internal morphemic structure, its behavior with respect to Celtic mutations, and its incorporation of object clitics and old sentence-initial particles into a single "univerbated" word --- can be plausibly traced back to Indo-European. Others have disagreed with Watkins in matters of detail (thus Meid 1963, Boling 1972, Cowgill 1975, McCone 1979, inter alia) --- but the general strategy stands. The clear implication is that this, as opposed to an overenthusiastic appeal to

substrates, is the way to do responsible historical linguistics.²⁴

Yet such an approach overlooks three fundamental points. The first is simply the fact that all the various developments which Celtic underwent, even if individually derivable from Indo-European (and not all are), did nonetheless all occur within the selfsame language group, and were carried much further here than elsewhere in Indo-European. Why? The second point is that the resulting language type is indeed --- impressionistically --- radically divergent from Indo-European, but in many ways shows typological convergence with Hamito-Semitic. Again, why?

The third consideration, a subtler point, is that genetic and substratal explanation are not mutually exclusive. For a given feature, there is nothing contradictory about tracing its purely genetic filiation, and at the same time postulating that that particular line of development (out of all the other possible lines of development from the protolanguage) was followed because of contact or substratal pressure.²⁵ To the contrary, the two approaches complement each other: the genetic filiation can provide a plausible "how", the substratal explanation a plausible "why". Assuming for the sake of argument a substratal explanation, it is surely at least as natural for substratal influence to have nudged Insular Celtic in the direction of an option preexisting in early Celtic, as for the substrate to have "infected" Celtic with a feature utterly alien to it.

²⁴ This aversion to substratal explanation, where a minority of beleaguered substratalists face hostility or indifference on the part of the majority, mirrors in small a longstanding trend throughout most of Indo-European study. For European substratalism, see Craddock 1969.

²⁵ The point goes back at least as far as Brugmann (sec. 2.3.1.2).

We will return to this point in the next chapter (sec. 2.3.1.2). But it will be useful here to illustrate with a concrete example. Celtic came to favor VSO word order during the course of its development, as a comparison of Continental Celtic and Insular Celtic makes abundantly clear. Watkins (1963) traces this process in detail, showing how the standard Old Irish sentence-initial verb, schematically

P (E) V ... (= Particle (Enclitic) Verb),

developed from an earlier structure

P (E) ... V

through the process of fronting and univerbation of an original sentence-final verb. Going beyond this account of "how", Watkins further tries to make the case that the univerbation was all but inevitable --- that is, to pose the question "why?" and answer it in a purely language-internal way. A critical phase of the argumentation reads as follows (p. 40):

The univerbation involved simply the movement of V as far forward in the sentence as it could go, up to E.... The [object] enclitic E was immobile, in second position in the sentence; it was phonologically and syntactically bound.... Irish could have developed free and mobile forms for its pronominal objects, as did other languages. But it did not do so [italics mine], and this left no choice as to the direction and manner of the process of univerbation.

The point I wish to emphasize here is not the details of the argument but its overall structuring. The italicized statement is the bottom line in this explanation, and Watkins simply asserts it. Thus he

answers the "why" of VSO only by invoking another unanswered why. But the latter is just as puzzling as the former. Why should Celtic fly in the face of a general Indo-European "drift" toward independent object pronouns?²⁶ Here an appeal to substratal influence, if valid, could provide important insight. Arabic, Egyptian, and Berber all have VSO order and object coding on the verb,²⁷ exactly the right structural configuration. If we had reason to trust appeal to substrate as a possible mode of explanation, we might have a much stronger "why" than Watkins can provide. --- None of this is intended to prejudge the issue. Rather, the point is simply that a genetic and a substratal account need not be mutually exclusive.

1.5.5 Absence of lexical resemblances

A quite distinct line of attack on the substratal theory is that it posits structural influence with no trace of lexical influence.²⁸ Were the syntactic changes in Celtic alleged to have occurred by contact-induced structural borrowing, the criticism would seem fairly serious, for structural borrowing invariably is preceded by lexical borrowing (Thomason and Kaufman 1988:37, 113). In the case of substratal influence, however, language shift is involved. Here the issue of lexical

²⁶ Indeed, later forms of Celtic do develop object pronouns.

²⁷ In Egyptian the object pronoun is a clitic which usually occurs adjacent to the verb.

²⁸ This criticism may or may not apply to the areal approach. Nothing beyond structural resemblance is assumed explicitly; still, linguistic areas normally do display areal diffusion of vocabulary as well as structure. On the other hand, the "area" Wagner proposes involves such enormous distances that perhaps only minimal penetration of vocabulary might be expected.

penetration is not theoretically clear. It would be expected that a substratal people seeking to learn a new language would focus first and foremost on acquiring its most obvious component, namely vocabulary; there would be little motivation to retain words from the substratal language (1988:39). Thus we might expect little substratal influence on vocabulary. Yet there are cases of substratal structural influence accompanied by lexical influence, such as that of Cushitic on Ethiopic Semitic (1988:117, 133-34; also Leslau 1945:79-81). Hence it seems prudent to keep an open mind on the matter of lexical influence of a Hamito-Semitic substratal language on Insular Celtic, and explore it as a real possibility.

In point of fact, the two language groups have no evident shared vocabulary; if attempts have been made to ferret out this or that Celtic word of putative Hamito-Semitic origin,²⁹ the scattered lexical look-alikes do not rise above the noise level of coincidence. As suggested, this is perhaps what we might expect from a substratal scenario. But even if we allow the theoretical possibility of vocabulary borrowing, there are two reasons that any such instances might be exceedingly difficult to detect.

First, the date of such putative borrowings is unknowable, as is the plausible (close) genetic affiliation of the source language, the language of the migrants from North Africa. The latter is unlikely to be identical to any actually attested Hamito-Semitic dialect. With luck

²⁹ Thus for example G. B. Adams (1956:15-16, 1975:240), who suggests a link between Irish tulach and Semitic tel "hill", running the connection via Sardinian. On the general problem of pre-Indo-European etyma in Celtic, see Campanile 1976, with no mention of Hamito-Semitic.

it will have been closely related to a known Hamito-Semitic language, a probable candidate being perhaps Numidian, the presumed ancestor of modern-day Berber. Numidian, however, is attested only through a few, largely unvoiced, inscriptions from Punic times. Rössler (1958:120) does point out the striking similarity between Berber and (his rendition of) Numidian; thus in default of Numidian we might perhaps appeal to modern-day Berber, though the fact of 2500-odd years of intervening history is not an encouraging factor as regards preservation of vocabulary. Ranging farther afield, Egyptian and Semitic are well attested at early periods of their history, but here again sources are unvoiced; the single exception, Akkadian (written syllabically), is geographically furthest from the Mediterranean locus of the posited substratal language, and structurally furthest from the Celtic/Hamito-Semitic type.

A second source of difficulty involves the particular phonological facts of the two language groups. On the one hand, the Hamito-Semitic languages are rich in laryngeal consonants (h, ʔ, ^c, ʕ) --- and we know, from extensive data on loanwords from Semitic, that these sounds are liable to distortion or elimination when words containing them are borrowed into other languages. This in itself could disguise a particular instance of borrowing to the point of unrecognizability, especially in light of the absence of phonemic /h/ and the inconsistencies regarding graphemic <h> in the oldest attested Insular Celtic language, Old Irish. On the other hand, the Insular Celtic languages themselves are notorious for the degree of phonological alteration they have undergone, featuring wholesale vowel syncope, changes in vowel quality, and (at least at an early date) phonologically conditioned mutations. This, too, would constitute an effective disguise. In brief, even if lexical borrowing did

occur, we cannot do better than guess at the identity of a plausible source language; the lexicon of this language (or even of a plausible cousin language) may be known fragmentarily at best; and the particular phonological structures and histories of the language groups in question would conspire to render such cases undetectable.

Thus we certainly should not dismiss the case for prehistoric contact simply on the basis of lack of lexical borrowings. Given the circumstances, we could scarcely expect to find any.³⁰ And given a substratal scenario, none need be expected to have occurred in the first place.

1.6 The typological method

The substratum scenario is at least *prima facie* plausible. The structural similarities between the two families have indeed struck many linguists (including myself) as remarkable; Pokorny's (and Wagner's) whole approach is predicated on the (to him self-evident) assumption that such remarkable syntactic parallelism "kaum ein Zufall sein wird" (1949:245), but could only be an indication of contact. Indeed, much the same reasoning is often applied when two languages show strong, systemic, and idiosyncratically "quirky" morphological resemblance; this is in fact one of the standard arguments, not merely for contact, but for the stronger link of genetic relatedness. There is an important

³⁰ Adams 1980 goes into considerable detail on the methodological difficulties attendant on such borrowings --- significantly, without adducing any concrete examples, which "in the present state of our knowledge would be to demand too much" (p. 63). See further discussion in sec. 2.5.1.

theoretical question here: just when is a syntactic parallel to count as strong, systemic, and quirky? For syntax far more than for morphology, this question is inherently a typological one. Because a morphological parallelism may encompass both the patterning of morphs and their phonological forms, and because the parallelism may be either paradigmatic or syntagmatic or both, one may come away with an overpowering conviction of quirky systemic similarity just by inspection. But syntax, being pure pattern with no analogue to "phonological form", and lacking much in the way of an interesting paradigmatic dimension, lends itself far less to an intuitive grasp of "quirky" similarities between two languages or language groups. Rather, the only way to determine what counts as unusual is to find out, the empirical way: by going on a broad crosslinguistic excursion and seeing what actually happens. And this falls squarely in the domain of typology.

1.6.1 The concept of typology

Typology is the systematic study of cross-linguistic similarities with several interlinked aims in mind:

- 1) To uncover and delineate interesting dimensions (= variables) of cross-linguistic variation;
- 2) To determine what values can be assumed along such dimensions;
- 3) To determine the distribution of values along a dimension: which values are more common than others?
- 4) To uncover correlations among dimensions which may at first have seemed "orthogonal".

The use of the term "dimension" here is meant to suggest loosely the metaphor of an n-dimensional Cartesian space, each axis corresponding to a different typological variable. The identification of a potentially "interesting" dimension is a creative and highly subjective process; the number of such dimensions is boundless, and there are no fixed answers. In the context of the present problem, the typological dimensions will be roughly those defined by the similarity features observed between Celtic and Hamito-Semitic. Indeed, I will be using the terms "feature" and "dimension" as near synonyms.

The most heavily studied typological dimensions in recent decades have involved word order, thanks to the work of Joseph Greenberg (1966a) and others; and I will illustrate the above 4 points within this by-now familiar domain. There is, first of all, a dimension "Basic Word Order" specifying possible permutations of Subject, Verb, and Object (S, V, O) (Point 1). A language will assume one of six values along this dimension:³¹ SVO, SOV, VSO, VOS, OVS, OSV (Point 2). Of these, SVO and SOV are the commonest, with VSO a significant minority, VOS a very small minority, and OVS and OSV attested only in a handful of languages of South America: schematically,

SVO,SOV > VSO > VOS > OVS,OSV (Point 3).

These possibilities can be laid out as a histogram (Figure 1).³²

Finally, Point 4 necessarily requires reference to another typological

³¹ Properly speaking, we should have a 7th value "not applicable", for languages having no basic word order. This complication will be of immense practical importance in evaluating the typological data to be presented in the second half of this study; see sec. 5.1.3.

³² The graph is not to scale; it is intended to be merely schematic.

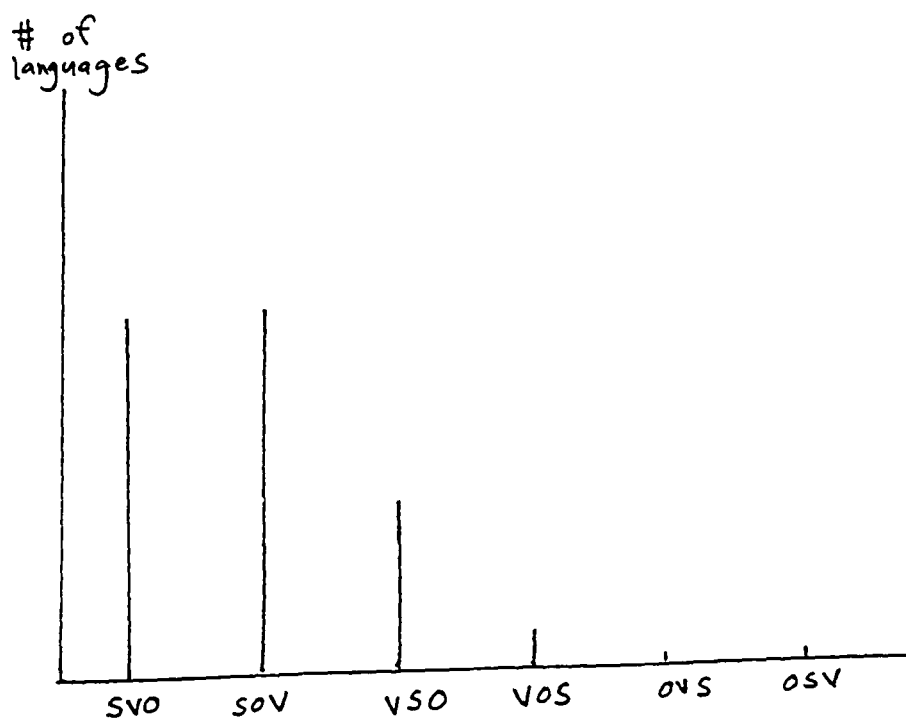


Figure 1
(Basic word order type)

dimension --- for example, the dimension specifying possible orders of Noun and modifying Genitive (N-Gen, Gen-N), or that specifying orders of Noun and Adjective (N-Adj, Adj-N), or that specifying prepositionality vs. postpositionality (Prep-N, N-Postp). As Greenberg and others have shown, these are not independent of the dimension "Basic Word Order", but show correlations and implicational universals of various kinds; for example, the value "VSO" overwhelmingly implies the values "N-Adj", "N-Gen", and "Prepositionality" along the appropriate dimensional axes, a constellation which may be termed the "ideal VO macrotype".

When particular values along different dimensions show an affinity of this sort for one another, the ensemble of these values may be termed a type. A major goal of typology is thus to uncover and investigate linguistic types. Since Greenberg's work, word order has so dominated typological study as to become the unmarked sense of "typology". There is nothing logically necessary about this, of course. Any typological dimension(s) can be the basis of a linguistic type. Earlier generations of typologists focused on the basic morphological make-up of words (with isolating, agglutinating, synthetic, and polysynthetic types). In recent years a major typological division of languages into Head-Marking and Dependent-Marking types has been proposed (Nichols 1986); the issue is whether the marker of a grammatical dependency occurs on the head or the dependent of a construction. One concern of the present study is whether the constellation of features shared by Celtic and Hamito-Semitic, or some subset thereof, might not form a type by virtue of clustering together. (Note that these features do not primarily concern word order; word order is only one item in the list of commonalities presented in section 1.2.)

1.6.2 Typological features and their values

In asking about the typological distribution of a linguistic phenomenon P shared by two languages, the question makes sense only insofar as the phenomenon is taken as one possible realization of a broader paradigmatic class U (= Universe) --- in other words, as one possible feature-value F1 assumed by a feature F, the latter conceptualized as a dimensional axis U. The problem is: given the phenomenon P, what are we to select as this universe U, and what are to be the possible values of U in addition to F1? The answer may seem obvious in particular cases, as in the case of word order. If F1 = "VSO", then the "right" assignment of U (we feel intuitively) is the dimension comprising "all permutations of V, S, O", including a total of 6 possible values.³³ However, things are not always so simple. Supposing (as in the CHS languages) that two languages share the feature-value F1 "having syntactically significant word-initial change". A reasonable candidate for U would seem to be the two-member class {"having such change", "lacking such change"}. However, there are also languages with syntactically significant word-final change. Do they represent a third value of a differently conceived universe U? Further, in some such languages word-initial change encodes only a single grammatical distinction, whereas in others it encodes a hodgepodge of different distinctions. Shall these different possibilities be subsumed as different values F(i) in the again-revised universe U? The point is that the determination of

³³ This can be questioned, however. Some researchers have preferred to deal in a two-valued dimension, "VO vs. OV". Others point out that the sentence contains additional material besides S, O, V, so that one should distinguish orders SXOV vs. SOXV, etc. And even the very concepts of Subject and Object can be called into question.

a "universe" dimension U, an absolutely essential step in doing typology, is often unavoidably subjective. This can be a major source of imprecision in determining how features are typologically distributed. We will face this issue on a case-by-case basis (with less precision than could be desired) when we consider the individual features in Chapter 5, and as a general theoretical problem in sec. 5.1.1.

Another problem is that many typological dimensions logically fall into a hierarchy. The dimension "having contour tones or not" makes no sense unless the language is in fact a tone language --- i.e., unless it assumes a particular value ("yes") along the dimension "having tone". Some languages, then, will simply be irrelevant to a global investigation of contour tones. To bring the discussion back to Celtic and Hamito-Semitic, we will need to examine (for example) the typological distribution of conjugated prepositions; the question cannot even be entertained unless the language has prepositions (or adpositions) in the first place. Similarly with the genitival construction "X the-Y". Typological data on the positioning and the number of occurrences of the article can be gotten only from languages that have articles at all; and even for such languages, the choice of a "universe" dimension U is far from obvious. This problem, and others of the same sort, will confront us again in Chapter 5.

1.6.3 Skewed distributions and typological inference

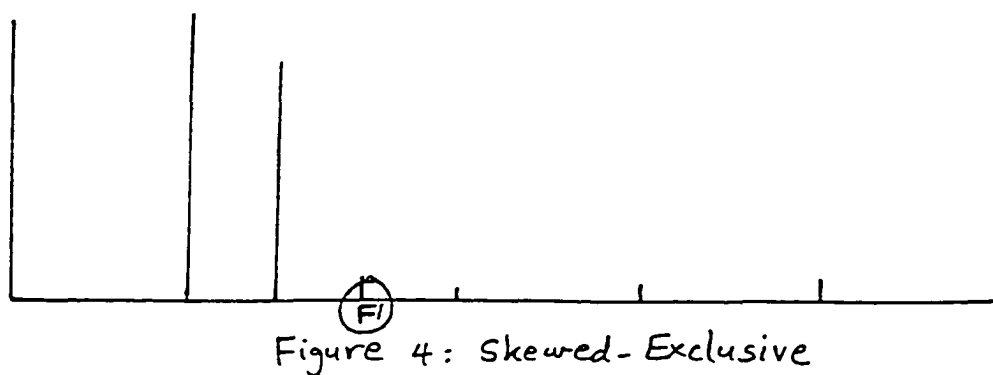
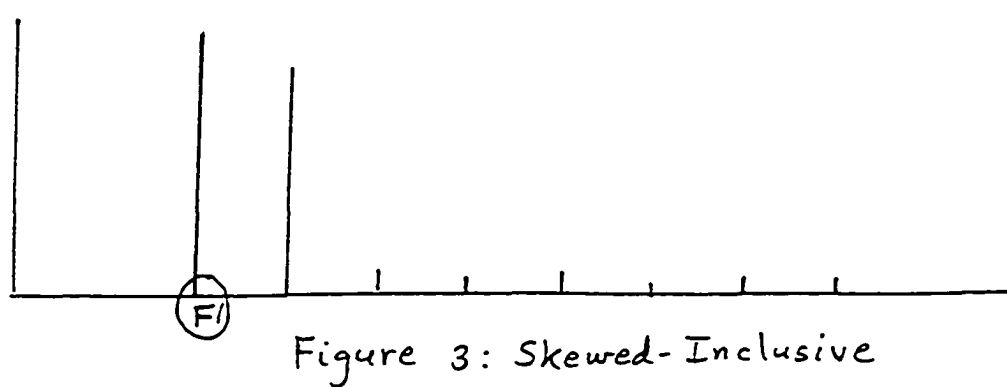
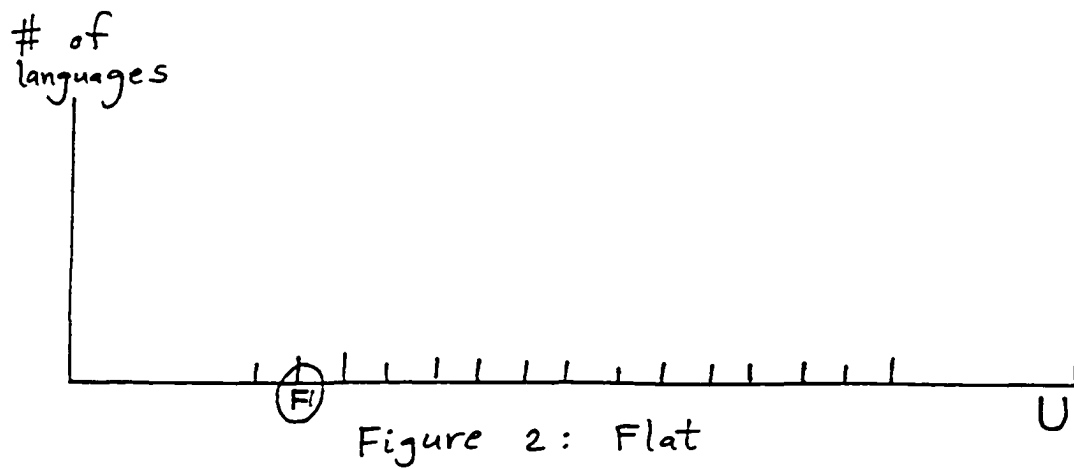
We will henceforth assume that every feature-value F1 has been assigned (however subjectively) a "universe" dimension U, and proceed to examine as a theoretical problem the different types of distributions

that a typological feature can have. Distributions will be displayed, as before, using histograms, with the horizontal axis (the "universe" dimension) running through the possibilities inherent in U, and the vertical axis giving their distributional frequencies of occurrence (in number of languages). In the context of the present work, we are interested in cases where two languages or language groups (in fact the CHS languages) have some feature-value F1 in common. The circled (F1) on the graphs will represent this particular feature-value of U, the value held in common by Celtic and Hamito-Semitic.

There are three extreme distributional types that will be of crucial importance as limiting cases; we may label these flat, skewed-inclusive, and skewed-exclusive, as indicated in Figures 2, 3, 4.³⁴ Looking at these three distribution types in turn, I will ask the question: supposing two languages share the circled feature-value of some feature F; what is the most likely explanation of the commonality?

In a flat distribution, all values of U are more or less equally (un)likely, as would be the case with repeated tosses of an unbiased coin or dice. Such a distribution applies, for example, to the feature "number of actants coded on the verb". There are no robust trends, no strongly preferred values; all four values of the feature (viz., 0, 1,

³⁴ As used in this study, the term "skewing" diverges from its technical mathematical sense, which refers to the degree of distortion in the shape of a bell curve such that one side of the bell slopes more steeply than the other. The notion of peakedness or spikiness I am referring to here goes by the technical name of "kurtosis". I retain the term "skewing" as the most accessible to a non-mathematical audience; such phrases as "degree of skewing" or "highly skewed", which I will be using frequently, would sound distinctly odd if rendered as, e.g., "degree of peakiness" or "highly spiked".



2, 3+ actants) are empirically observed to occur in roughly the same number of languages (with a minor preference for 2 actants). With such a feature, if two languages share the same value F1 of U, it is simply a random result. In such a case we can say, on a principled basis, that there is no good reason to seek any explanation for the resemblance other than coincidence.

In skewed distributions, on the other hand, certain values of U are observed empirically to be far more likely than others, as in repeated throws of weighted dice. The distribution shows a spike. In this case, the type of inference to be drawn will differ radically depending on whether the set of languages realizing the feature-value F1 represents a majority or a minority subpopulation. If F1 represents a majority type ("skewed-inclusive"), there is again no reason not to appeal to coincidence. However, if F1 represents a small minority type over against a dominant majority ("skewed-exclusive"), then F1 can properly be seen as something exceptional, as "bucking a trend". I will call features of this latter kind exotics. In the realm of basic word order, for example, the object-first types (OSV, OVS) are extremely rare vis-a-vis the dominant majority types; in the realm of click vs. non-click languages, the click languages form a tiny minority. It is when two languages show an agreement in an exotic feature that we feel most strongly that something is likely to need explaining. The case is analogous to throwing two weighted dice and having the same value come up twice, but not the value for which the dice are biased! Here coincidence seems maximally unlikely (though always possible), and the case for a historically framed account (migration, contact, substratum) becomes about as good as it possibly could be. Nothing is proven, of course; but it is highly

suspicious that two languages should independently be bucking the trend, and in the same way. If a plausible geographical and historical case for contact-or-relatedness can be made at all for the two languages, then the two occurrences will in effect have been reduced to one. In the above examples, it is significant that all OSV and OVS languages occur in or near the Amazon basin in South America, and that all click languages occur in Africa (and almost all in southern Africa). Given the overwhelmingly skewed distribution, and the plain fact of areality available as a ready explanation, it would seem perverse to insist that the co-occurrence is after all just a coincidence. The same principle will apply, *mutatis mutandis*, even if the fact of contact is less patent --- as in the case of the similarities between Celtic and Hamito-Semitic, where indeed many of the occurring feature-values turn out to be the minority type in a skewed distribution. Here again (unlike the flat distribution) there is something calling for explanation --- and an account framed in terms of prehistoric contact should therefore be viewed as favorably as possible. The more drastic the skewing, the more attractive such a posited historical explanation becomes.

We turn now to the case of several interacting dimensions or features along which two language groups show agreement --- for example, the three dimensions N-Adj vs. Adj-N order, G-Gen vs. Gen-N order, and Preposition vs. Postposition. The considerations in the preceding paragraph, of course, will still apply to each feature separately and to all the features collectively. In particular, if multiple features show a skewed-exclusive distribution, the case for rejecting coincidence as an explanation becomes well-nigh irresistible: to continue the earlier metaphor, multiple throws of a pair of weighted dice keep coming up in

agreement, but over and over the value that surfaces is not the favored one. That the favored value repeatedly fails to show up is itself remarkable; that the unfavored values which do show up repeatedly agree with each other is astonishing.

On the other hand, it now becomes crucial to determine whether all or some of the various features are actually independent of one another. If they turn out not to be independent but are intercorrelated, recurring as a bundle in languages from all over the world, then a third mode of explanation becomes likely: natural typological affinity among the features (as seen, e.g., in the "ideal VO macrotype" [sec. 1.6.1]). The particular co-occurrence of features in Celtic and HS would then be seen as natural; on the basis of the identical co-occurrence pattern in many other languages, we would expect them to occur in a bundle here as well (though, to be sure, we would still not know why they should so co-occur). What had been thought to be a remarkably large number of similarities would now no longer be as remarkable or as large, for (some of) the various features would no longer count as independent variables but rather be linked manifestations of a single "macrofeature". The occurrence of this "macrofeature" as a whole might then be a coincidence, but the co-occurrence of the subfeatures comprising the macrofeature would be properly taken as a natural typological possibility of human language.

The typological method thus exposes the choice of "contact vs. coincidence" as a false dichotomy. Even if the set of resemblances truly is non-coincidental, there is a third choice: natural typological affinity. Depending on the empirically observed distribution of the

features in question, one or another of these three options will be the most plausible. In fact, as remarked, an explanation based on contact will turn out to be the most plausible; but a priori any of the three options could have represented the favored explanation.

1.6.4 The survey

The typological survey to be carried out in the second half of this study has as its goal to examine grammars of languages throughout the world, determine thereby the empirical global distribution of the CHS features, and argue from the observed distribution to the most likely explanation for the similarities. As indicated in sec. 1.1, there are four explanatory possibilities to choose from, conceptually split into historical and ahistorical:

Historical:

- (a) Genetic relatedness
- (b) Contact (nongenetic historical link: areality, substrata, etc.)

Ahistorical:

- (c) Typological affinity (correlation) of the features
- (d) Coincidence

Since genetic relatedness is excluded, we must choose among a triad of options. And the choices, as we will see, need not be mutually exclusive. There is, however, also a fifth mode of explanation, or rather meta-nonexplanation: radical skepticism. This is a willed abstention and agnosticism, a commitment to treating the CHS problem as one intrinsically not amenable to solution. Superficially this may seem

identical to an appeal to coincidence; but the radical skeptic does not want to go beyond coincidence. To those who hold this attitude, any link between Celtic and Mediterranean Hamito-Semitic has the apodictic absurdity of the flat-earth theory.

It is unclear what, if anything, could convince the radical skeptic. But the global distribution of the features, whatever it may turn out to be, is a matter of fact, and different distributions automatically carry with them different explanatory implications. If the features tend to show a flat distribution, coincidence has the best explanatory claim --- not out of skepticism but in deference to fact. If many of the features are exotics, conversely, coincidence is least likely. If the features turn out to be strongly correlated, an explanation in terms of typological affinity is likeliest. If little correlation is observed, a historical explanation becomes likeliest simply by process of elimination. If the CHS type recurs in Siberia or Bolivia, then clearly the type can arise independently in unconnected contexts, and the attractiveness of a historical explanation linking Britain and North Africa fades. Conversely, if the CHS type shows up nowhere else, then the argument in favor of prehistoric contact gains in plausibility. And if a weakened form of the CHS type is found to exist over a broad geographical zone, then the problem will acquire an areal aspect which may help contribute to an explanation. These and other similar arguments thus provide a nonimpressionistic way of gauging the likeliest solution to the problem.

Chapter 2: The history of the problem

2.1 Beginnings

Something about the ancient Near East has always drawn the Celtic imagination. The Auraicept na n-Éces (Scholars' Primer), perhaps the earliest attested Irish compilation of grammatical and poetic lore,¹ situates the origin of the Irish language and people at the Tower of Babel (Calder 1917:9ff.). Similarly, the Lebor Gabála Éirenn (Book of the Taking of Ireland; earliest ms 12th century [Macalister 1938(I):xi]) portrays Nél, ancestor of the Goidels, conversing with the Israelites at the Red Sea centuries before the coming of the Goidels to Ireland (II:59). In Wales, Geoffrey of Monmouth's 12th-century Historia Regum Britanniae, which rooted the Welsh in Trojan antiquity, was read eagerly and translated into Welsh as the Brut y Brenhinedd. Later, in the 16th to 18th centuries, scholars sought to give the Welsh a pedigree going back instead to the Biblical Gomer --- etymologically "identical" to Cymru! (Pezron 1706) --- or to the Phoenicians (Sammes 1676, following Samuel Bochart); Vallancey (1786) similarly derived Irish from Phoenician.² This turn to the Orient embodied a craving for an ancient and

¹ Though Middle Irish in language, the work is clearly based on Old Irish sources (Calder 1917:xxiii). For a keen evaluation of the Auraicept see McManus's 1991 book on ogam (passim, esp. 137ff., 147ff.).

² See e.g. Adams 1956:8-9 for further 18th- and 19th-century attempts to forge pseudohistorical links between Irish and the Mediterranean. Even in the 17th century, this sort of orientalism was by no means

glorious lineage, even one predating and thus surpassing the antiquity of classical Greece and Rome --- a movement sometimes dubbed "Celtomania".

But it was only by stages that this quest for historical roots in the Near East came to include a hunt for linguistic roots. In the 7th-9th centuries, Irish monks were in the forefront of European Bible study, and their exegesis frequently sought to illuminate a Biblical word or phrase by presenting it in the "tres linguae sacrae", Latin, Greek, and Hebrew³ (McNally 1958); yet they, like their great English contemporary Bede, seem to have had almost no working knowledge of Hebrew (Smalley 1952:35-36, McNally 1958:397), even inventing Hebrew words to enliven their commentaries (Herren 1979:61).⁴ The writer/redactor of the Auraicept na n-Éces again acknowledges Hebrew as one of the three "principal languages" (Calder 1917:13), even ascribing Hebrew origins to Fenius Farsaidh, the inventor of the Irish language (79). Had he had the slightest glimmering of a special Irish-Hebrew affinity, surely here, in the origin myth, would have been the place to use it to advantage; yet the author's boast, that Irish was created by selecting "what was best, widest, and finest of every language" (81), gives Hebrew no pride of place among these 72-odd tongues.⁵ In the universally accepted; see Morgan 1973/74.

³ The three languages of the inscription on the Cross (John 19:19-20).

⁴ For knowledge of Hebrew in the early Middle Ages, see Thiel 1973; for Hebrew studies in England, see Jones 1983. Neither work appears to address the question of Hebrew in Ireland.

⁵ The above ideas draw on the encyclopedic twenty-book Etymologiae of the 6/7th-century bishop Isidore of Seville (see Calder 1917:xxxi ff.), which enjoyed immense popularity in medieval Europe and in Irish circles in particular (McNally 1958:396). Such speculation on the origin, differentiation, and relative status of post-Babel languages was very popular in the Middle Ages (see e.g. Wolff 1971:110-13).

Lebor Gabála Éirenn, Hebrew (catalogued miscellaneously between Albanian and Arcadian) is again only one of 72 languages that went into the formation of Gaelic (II:13, 83, 89, 148ff.); elsewhere, in a poetic stanza (I:178-79), there appears a single isolated mention of the Hebrew name of God (YHWH), the only linguistic appeal to Hebrew in the compilation.⁶ To seek in such works any interest in, or awareness of, a linguistic affinity would be futile; obviously these writers did not envision any special tie with the language of their Near Eastern cousins.

There would seem to be one very early exception to this generalization. F. J. Byrne notes in passing that the sixth- or seventh-century grammarian Virgilius Maro, who is argued to have had strong connections to Ireland (see Herren 1979), "long ago remarked the similarity of Irish to Hebrew in placing the verb at the head of the sentence" (1962:199) --- a remark picked up by Pokorny (1964:76) and echoed by Greene (1966b:43). This comment perpetuates a Celticist tradition, by now apparently part of the established lore, ascribing to Maro explicit mention of the verb-first word order of Irish (thus Herren 1979:56-57, citing Mac Cana 1973:93-94, Meyer 1913b, and others). But Maro's thinking

⁶ It is noteworthy that Macalister's restituted text (I:178) presents the Hebrew letters of the Tetragrammaton backwards: "iae, uau, iae & iath" = HWHY (compare the correct Hebrew spelling of the letter-names, "iod he uau he" = YHWH). The reversal is ascribed (I:263) to the poet's unfamiliarity with the right-to-left order of Hebrew writing. If true, this would bespeak a total linguistic ignorance of Hebrew. G. B. Adams (1956:8) accepts the reversal but offers a different explanation: in context the letters comprise an incantation recited by the Devil, and as such might have been reversed deliberately. In fact, however, the reversal itself is problematical. The individual ms readings as given in Macalister's apparatus (I:178) present a garbled picture. Most of the mss have "ia ae uau & aiath", which at best might be a distortion of "he he uau & yod" = HHWY but cannot plausibly represent either YHWH or HWHY. Even with the other variants, no one "correct" reconstruction suggests itself.

and language at their best are seldom clear, and in this instance the confusion is compounded by a textual crux. The relevant passage, and the basis of the tradition, is a clause at the beginning of chapter 5 ("De Nomine") of Maro's Epitomae: "cum in hi bonorum/hiborum elocutione et conpositione primatum estimatur uerbum" (Huemer 1886:26) --- that is, the verb is considered "primatum" (primary). There are two problems here: the intended sense of "primatum," and the textual disagreement between the two manuscripts regarding the underlined word(s).

As to the latter difficulty, neither version makes clear sense. Meyer says "we should undoubtedly read Hibernorum" (1913b:24, cf. p. 8), but others have just as confidently suggested "Hebreorum" (per Tardi 1928:147, note 32); taking "bonorum" at face value, Tardi translates "ceux qui s'expriment bien" (1928:60-61), while Herren proposes "of the good men of Hy [= Iona]", i.e. the Irish (1979:56). Regarding the notion of the verb's "primacy," Meyer opts for the sense of linear precedence, asserting that Maro's chapter 5 "treats of the order of words in the Latin sentence" (1913b:8, likewise Mac Cana 1973:93, Herren 1979:56). This view requires that this particular passage be construed as referring to some verb-first language (Irish); for to assert verb-first word order for Latin would be absurd. But neither the actual chapter text itself (almost entirely devoted to such time-honored morphological topics as gender, number, case, proper vs. common nouns, and word families, none having the remotest connection with word order) nor indeed the theory and praxis of medieval grammatical tradition (which never paid much attention to word order) seem to encourage such a verb-first interpretation. Undoubtedly Maro's puzzling text might be taken as a statement about verb-first word order, and might be referring to

Hebrew or to "Hibernic" (though hardly both at once, contra Byrne); but it appears much more likely in context to be invoking a sense of the abstract conceptual primacy of the verb, as the pivot and prime mover in clause-level syntax. This is implicit in Tardi's translation (1928:60-61): "Puisque, dans le langage et la construction de ceux qui s'expriment bien [= Hi+bonorum], le verbe est regardé comme occupant la première place" --- that is, the verb has pride of place among the words of the sentence and is in this sense "primary".⁷ Such a preoccupation with the central essences of things is very much in the spirit of medieval grammar. And since this interpretation is a universal one, applying equally to all languages, the central motivation for an Irish-centered construal of the passage disappears.

The whole "verb-first" tradition, in sum, is merely an interpretation, and not a particularly well-founded one; and an explicit linkage between Irish and Hebrew has no foundation in Maro's text on any interpretation.⁸

⁷ See further the beginning of Maro's chapter 7 ["De Uerbo"], where the verb, as "principalem partem orationis" ("la principale des parties du discours"), is presented as the prime force binding together the pieces of the sentence --- "in positione quassorum principatum tenet" ("il [le verbe] occupe ... le rang essentiel dans la composition des phrases") (Tardi 1928:84-85, Huemer 1886:50). This statement occurs in a context which explicitly mentions Latin, precluding any verb-first interpretation.

⁸ For Maro's influence on native Irish grammatical tradition see Zimmer 1910:1031-98 --- with no mention of verb-first word order. For recent comprehensive discussion of Maro see Herren 1979, who tentatively endorses the verb-first interpretation of "primatum" but with little in the way of positive justification (56-57): "It is hard to see [w]hat else might be intended by the words. One could only give first place/pre-eminence to a part of speech in elocutione (utterance, pronunciation) by using it first in a phrase or sentence" (56; 'what' is misprinted as 'that'). As Herren argues, Maro has in mind here the order of presentation of the parts of speech as found in traditional grammatical treatises (cf. Priscian's ordo naturalis): Noun before Verb. Maro

By the 16th-18th centuries, by contrast, Celtic antiquarian scholars were regularly appealing to linguistic data to buttress their claims. Their standard method of argumentation was twofold. First came an appeal to authority: a sophisticated and often sophistic milking of every possible drop of historical data from the Bible, Josephus, Herodotus, Caesar, etc., reinterpreted to suit the author's argument. Coupled with this, especially as regards Hebrew/Phoenician, was an etymological argument: a phonic and/or orthographic comparison between Celtic and queerly transcribed Hebrew, aimed at the establishment of etymological equations between words and/or names. Such etymologizing could be far more dramatic: occasionally we find entire blocks of text in Hebrew or Phoenician interpreted, sound for sound, as literally being Celtic. Thus Edwards (1936[1677]:405-21) has sixteen pages (!) of Hebrew Bible passages presented as Welsh; Vallancey (1786:306-24) renders the Punic passages in Plautus's Poenulus as Irish text.⁹ Accompanying all this might be an elaborate sociolinguistic scenario for the language break-up attendant on the Tower of Babel catastrophe. There were glimmerings of an awareness of the possibility of coincidental resemblance, and of the temptation to make the data fit a clever analyst's agenda (cf. Sammes 1676:43), but this did not diminish the zeal of these early etymologizers: the word lists can run into the hundreds (e.g. Edwards

instead urges the reverse order; his justification, Herren suggests, is the fact of verb-first word order in Irish. Yet Priscian explains and rationalizes his "order" in terms of an aprioristic logico-philosophical hierarchy, not in terms of word order (Covington 1984:6); plausibly, Maro would have done so as well. For Maro to appeal to word order, and in a vernacular language (Irish) at that, would flout tradition twice.

⁹ The procedure is exactly the same as that in the spoof Mots d'heures: gousses, rames ["Mother Goose Rhymes"], whereby the actual syllables of the original text are made to make a kind of sense in a different language.

1936[1677]:394-405, Rowlands 1723:278-86).¹⁰

Only rarely in these works is there any reference to structural similarities between Welsh and Hebrew. Yet it seems utterly natural that Welsh scholars who learned Hebrew should have been struck by such resemblances, and for one man, at least, they were of central importance. John Davies, a scholar of Hebrew as well as Welsh, states in the preface to his monumental Welsh-Latin dictionary of 1632 (passage cited and translated from the Latin by Williams, 1973/74:208):

[The Welsh] language seems to me ... to be of a genius so different from all the European and Western languages, at least such as they are at present and have been for many years past, that there is not even the slenderest foundation for thinking that it might be derived from them. And I am best pleased with their sentiment, who deem it to have taken its rise from Babel. It is my opinion that it is one of the oriental mother-tongues, or at least immediately sprung from these ... I may venture to affirm that the British tongue has an evident agreement and affinity with the Oriental languages in its words, phrases, composition or structure, and pronunciation of its letters.

This is no mere programmatic statement. A decade earlier, in 1621, Davies had published in Latin a superb Welsh grammar (illustrated with textual examples) abounding in structural comparisons to Hebrew.

¹⁰ Rowlands (1723:287) explicitly acknowledges his debt to Edwards's earlier compilation.

Neither language, for instance, has a true present tense: Hebrew uses a participle, Welsh a periphrasis with yn + verbal-noun (76-77). Both Hebrew and Welsh have N-Adj order (163). In both languages (171-72), relative clauses feature "ellipsis" of the relative element, and make use of "pleonastic relatives" (resumptive pronouns) --- as in

perish (the) day [I-was-born in-it]

"Perish the day on which I was born" (Job 3:3) .

Prepositions in both languages take affixal objects (180), which are "cum vocibus alijs in unam dictionem coalescentia" ["fused with other words into a single vocable"] (Preface). In the genitive syntagm "the X of the Y", only the embedded element "Y" can take the article (197).

In his enthusiasm to trace the "Oriental" origins of Welsh, Davies ekes out every possible parallel; many more points come up, some trivial and some erroneous (including the usual chance resemblances in word-forms, e.g. hi "she" in both languages [67]).¹¹ It could also be objected that many of Davies's parallels are illustrated with Bible passages, where the Hebrew original might be biasing the Welsh syntax; but this is more a theoretical than an actual objection, for the Welsh Bible translation is universally acclaimed a stylistic masterpiece.¹² In fact, the typological commonalities listed above would pass muster today with no difficulty. They represent the earliest concrete demonstration of Celtic/Hamito-Semitic structural parallels I have come across.¹³

¹¹ In Davies's dictionary, such etymological parallels to Hebrew (some very strained) appear on almost every page.

¹² Davies himself had a major hand in the translation (Morris-Jones 1913:v).

¹³ It is instructive to compare Davies's work with that of his grammarian predecessors. Gruffydd Robert's grammar (1567) makes frequent reference to Latin and Greek, but none to Hebrew. Henry Salesbury

Davies's degree of insight here appears to have been exceptional. More typical is the presentation found in Rowlands (1723). He, too, remarks in general terms "the near Affinity of our Phrase and Syntax with the most antient Hebrew Tongue" (39); he asserts that primitive British "took [its] rise from, and built [its] Structures upon, the Remains of that most antient Hebrew Tongue" (308; cf. 289, 317); he is also aware of phonological parallels (the alternation of b and v found in both Celtic and Hebrew) (276). But further concrete details are absent. Indeed, Rowlands's comparative word-list contains a highly revealing example of indifference to structural considerations:

Hebrew itho = Welsh iddo = "With him" [quoted verbatim]. For Rowlands the value of this entry (sandwiched in anonymously between "face" and "furnace") inheres purely and simply in the phonic resemblance between the two word-forms. That a perfect morphological isomorphism is also involved, with both words comprising a conjugated preposition with a 3-m-sg suffix, is passed over in silence.

By the 19th century, serious British antiquarians had grown thoroughly disillusioned with quasi-mythological tales of Celtic origins in the exotic East.¹⁴ There was something embarrassingly unserious about these orientalist accounts; scholars instead focused on archaeology, anthropology, philology, and history in more or less the

(1593) does mention Hebrew in passing (some 7-8 times), and in such a way as to suggest that he really knew the language. Yet his appeals to Hebrew seem almost incidental, and (with one exception, p. 57 [Syntax, Chapter 2]) have nothing to do with syntax. It seems reasonably clear, then, that Davies's observations of Welsh/Hebrew syntactic parallelisms do not merely restate a received grammatical tradition, but represent his own original contribution.

¹⁴ See Kendrick 1950 for an engaging and witty history of British antiquarianism.

modern sense of the words. Well before 1900 it was firmly established that Celtic was an Indo-European language, and any search for Celtic origins in the Near East henceforth seemed quixotic.

Yet the language and its speakers still posed something of an enigma. Celtic and the Celts seemed to embody a peculiar "dualism" (Zimmer 1910:1058), an odd intermingling of Indo-European and non-Indo-European elements. Zimmer (1910:1058-60) mentions the quinary counting system vis-a-vis the Indo-European decimal system, the rise of an agent nominal construction "man of VERB-ing" instead of older "VERB-er", and the reworking of the Indo-European terms for seasons of the year within a non-Indo-European calendric system; outside the realm of language he cites the institution of Mutterrecht (descent through the father's sister) vs. Indo-European Vaterrecht. Almost all scholars of the period also remarked on the short, darker-skinned physical anthropological type often observed among the Insular Celts, reminiscent of Mediterranean peoples (such as the Iberians) and contrasting with the tall, blond "Aryan" type which appears as the ideal in Celtic literature. The key to these oddities seemed to lie in the fact, known on archaeological grounds, that the Celts were not the aboriginal inhabitants of the British Isles; plausibly, the observed dualism represented a linguistic and cultural merger between the natives and the Indo-European invaders. The challenge was to identify this pre-Indo-European substratum. The presumably autochthonous Picts seemed a reasonable candidate, but too little was known of them or their language to provide much firm basis for serious research. Sundry links were sought between Celtic and other groups, notably ancient Iberian and modern Basque. There was nothing particularly linguistic about this quest: language data (such as was

available) was merely one of a number of factors which might contribute to a broadly anthropological investigation into prehistory.

Even when the investigation did take a linguistic turn, its anthropological/philological stamp is still unmistakable. Thus John Rhys (1890) singles out the common Irish naming formula "son of X", "servant of X", "hound of X", etc. as peculiarly non-Indo-European, and draws a parallel to Semitic onomastics (34) --- apparently the first to have done so (see Pokorny 1955:59). But for Rhys this linguistic detail is only an incidental curiosity, brought up in passing and disposed of in less than a sentence. Much of the article (40ff.), rather, has an anthropological slant, focusing on the "hound" names as a manifestation of totemism, something ostensibly alien to Indo-European (46).

2.2 John Morris-Jones

In 1900, in an appendix to Rhys and Brynmor-Jones's The Welsh People --- a book much concerned with Welsh origins, a question dear to John Rhys's heart --- the respected young Welsh scholar John Morris-Jones,¹⁵ a student of Rhys's, put forward a detailed comparative-syntactic argument suggesting a link between Celtic and "the so-called Hamitic languages of Northern Africa, including Old Egyptian and the Berber dialects" (1904/05:279).¹⁶ Poet, grammarian, and passionate

¹⁵ See Walters 1986 for a bibliography of Morris-Jones's works.

¹⁶ The appendix actually bears the date March 1899. Apparently M-J had a draft of this or a similar paper in hand by 1896, to judge from Kuno Meyer's allusion to it in a letter of that year (quoted in S. Zimmer 1986:281). His researches into the problem of Celtic-Near Eastern syntax go back at least as far as 1891, but the 1900 study was his first publication on the subject, as he himself tells us: "The substance of

advocate of the Welsh language, Morris-Jones was not one to shy away from controversy (James 1987:19). Several years earlier he had scandalized Welsh orthodoxy by exposing the supposedly ancient bardic institution of the Gorsedd as a recent counterfeit (James 1987:20ff.). And proposing a Celtic-Hamitic connection would surely raise hackles. Not only would it conjure up the ghost of the discredited orientalism of several centuries earlier, but it would do so by appeal to the hotly debated notion of mixed language --- Insular Celtic would be "Aryan in vocabulary, and largely non-Aryan in idiom" (1900:617). The late 19th century was the heyday of the neogrammarians; against the dominant tide stood the feisty figure of Hugo Schuchardt, with his opposition to neogrammarian algebraic purity and his vigorous advocacy of mixed languages and substratal explanation. A serious philological defense of a Celtic-North African connection would thus draw on a respected but minority current in contemporary linguistic thought. The idea may well have appealed to Morris-Jones.

Morris-Jones's focus is explicitly on "Hamitic" (Egyptian and Berber), not Semitic (1900:639, 1904/05:280). He did not embark on his investigations with particularly high hopes (280):

I confess I did not expect to discover much similarity between that part of the [present] paper which deals with Egyptian was communicated to Professor Rhys in April, 1891; the other comparisons were made later; but hitherto they have all remained unpublished" (1900:617). In fact, in 1891 M-J did publish something of tangential relevance (see Walters 1986:11): his encyclopedia article "Y Cymry" [The Welsh], referred to (apparently) in 1904/05:279 ("So I argued in 1891"). This study indeed links Insular Celts with putatively Hamitic-speaking "Iberians" (1891:102), chiefly on anthropological grounds --- but with no appeal to syntax.

the syntax of our pre-Aryan tongue as it persists in Celtic and that of its cognate languages in Africa. I hardly thought the chance of finding Welsh idioms in Renouf's Egyptian Grammar worth the seven shillings and sixpence asked for the book. I bought it, however, and found in it what I had not dared to expect.

What Morris-Jones found was a set of strictly syntactic resemblances extending over numerous spheres:¹⁷

1) Word order (619-20, 631): In Celtic, Egyptian, and Berber the dominant word order configurations are VSO, N-Adj, N-Gen --- the inverse of the patterns in "Aryan". Morris-Jones also points out a variety of devices (fronting, clefting) whereby, in all three families, the subject can occur initially; these parallels, however, are less exact.¹⁸

2) In all three groups, pronominal objects of prepositions are marked inflectionally, by a set "S" of suffixed pronouns fused onto the Prep ("conjugated Preps") (620-23, 633-34). This is completely "un-Aryan". The same set S has additional functions, varying from language to language:

-- noun possessor in Egyptian and Berber (but not, barring a few Prep-like exceptions, in Celtic);
 -- verb object in Irish and Berber (but not in Egyptian [contra M-J, whose source (Renouf) errs here]);

¹⁷ The following summary ignores marginal and/or dubious points. Page numbers refer to the 1900 article.

¹⁸ Recall in this connection Greenberg's word order Universal #6: All VSO languages allow SVO order (1966a:79).

-- verb subject in Egyptian and (in somewhat altered form) in Welsh, though M-J is less than explicit on the latter.

3) In Berber, when the verb is preceded by one of a closed set of preverbal elements ("a particle or a relative or interrogative pronoun" [634], and in particular a tense-particle [636]), the direct (and/or indirect) object pronoun is not suffixed to the verb but to this preverbal element. Morris-Jones's formulation errs in points of detail, but the phenomenon per se is basically as he describes it. Likewise in Celtic: in the presence of a preverbal particle, the object marker occurs as an "infix pronoun", sandwiched between the particle and the verb. There is nothing analogous in Egyptian; M-J (citing Brugsch) suggests as a possible Egyptian parallel that "the subjective suffix is attracted by negative ... particles" (635); but in fact this so-called "negative particle" is really a negative verb tm "to NOT", which of course takes subject suffixes like any verb (Gardiner 1957:262ff.).¹⁹ --- This construction is markedly un-"Aryan", though M-J does not say so explicitly.

4) In both Egyptian and Celtic, though not under identical circumstances, a verb with an explicit NP subject does not exhibit person/number concord but appears in a maximally unmarked form (621):

Welsh: daeth (3-sg)	"he came"	Egyptian: ^C nh-f	"he lives"
daethant	"they came"	^C nh-sn	"they live"
daeth y dynion	"the men came"	^C nh ntr	"the god lives"

The principle is transparent in Egyptian: verbs may take either

¹⁹ See sec. 2.4.1 below for further discussion.

person/number suffixes or full NP subjects but not both. Morris-Jones sees this system as having been overlaid onto the inherited Indo-European system: "In Welsh the idea of pronominal suffixes has been completely transferred to the Aryan inflexions of the verb" (622) --- an idea which must thus have been at odds with original "Aryan" practice. As the above table shows, however, the Egyptian/Celtic isomorphism is less perfect than would appear from M-J's discussion: the zero form accompanying a full-NP subject in Egyptian is truly devoid of personal marking, whereas its Celtic counterpart is simply the 3-sg (sometimes zero-marked and sometimes not, depending on the particular tense paradigm). Moreover, Morris-Jones's label for this unmarked form is unfortunate: he calls it an "impersonal", a not inappropriate term for the Egyptian but at cross-purposes with the quite distinct impersonal passive of Celtic. --- Nothing analogous exists in Berber, where subject concord is standard. To be sure, in subject relative clauses (including main-clause clefts) the verb appears in a "participial" form that is invariant in many dialects; Morris-Jones tries to capitalize on this by dubbing the Berber participle an "impersonal", but this RelCl usage (even with clefts) is obviously not on a par with the independent-clause phenomenon found with Celtic and Egyptian.

5) Egyptian and Celtic agree almost perfectly in making heavy use of periphrastic constructions of the type "preposition + verbal noun" ("he is at/in VERBing") (625). Egyptian also has a periphrastic construction of the form "do VERBing", with a good Welsh analogue (627). Morris-Jones says that Egyptian has periphrastic tenses involving a copula iw, which he compares to Welsh cleft constructions with sef and ys (626). However, the Egyptian construction is not cleft-like (Gardiner

1957:385); nor is it clear (despite Gardiner) that iw is truly a copula. If Berber has anything analogous, it is marginal. --- There is no discussion of this phenomenon vis-a-vis Indo-European.

6) Egyptian and Welsh both introduce a predicate in a copular sentence with a particle which is all but identical to the preposition "in" (the type "He is PTCL mighty", meaning "He is mighty") (628). Irish has an analogue in the type "He is in-his doctor" (meaning "He is a doctor"), but the construction "does not seem to be old in this form" (630); cf. Dillon 1927-28:324-26 for details. Berber too has a predicative particle (d) with similar function (637); here, however, the particle is homophonous to the Prep "with", though M-J mistakenly identifies it with "in".²⁰ In Indo-European such a usage occurs intermittently "after verbs of 'making'" (629), but in Welsh it is the unmarked type for predicates of all sorts.

The following two points are mentioned by Morris-Jones only as supplementary features. In neither case is a comparison made to Indo-European.

7) Egyptian and Celtic (Welsh and Modern Irish) both have relative clauses with a resumptive pronoun after prepositions, of the type "the man that you were talking with him" (637) --- though Morris-Jones notes that O'Donovan's Irish grammar dubs this construction substandard (in fact O'Donovan's prescriptivism here reflects the Old Irish rule

²⁰ Morris-Jones (637 and n. 2) seems here not to have properly understood his source (R. Basset 1886:49*-50* [glossary]; cf. p. 67 on the particle d-, and pp. 73-74 on prepositions).

[Thurneysen 1946:312-13]). Morris-Jones asserts this to be the Berber type as well, but he has misunderstood R. Basset's discussion (1886:20-21) and taken an Arabic-influenced construction as the Berber norm; Berber is quite different. See secs. 2.3.2 [13], 2.4.3 [8].

8) In all three language groups the copula can be omitted (637-38). Morris-Jones fails to note, however, how much more ordinary this is in Egyptian and Berber than in Celtic.²¹

In hindsight, despite errors of detail and a tendency to push the data a bit further than one should, Morris-Jones's achievement in this 25-page appendix is little short of amazing. Starting more or less from scratch, he has hit upon the core of the syntactic parallels between Celtic and Hamito-Semitic. The achievement is all the more admirable in light of the brevity and relative crudity of the early sources at his disposal: R. Basset's 1886 grammar of Kabyle (Berber) runs to a scant 88 pages, Renouf's 1875 Egyptian grammar to 66. Even more surprising, on two counts, is the modernity of his methodology: his comparison is both purely syntactic, explicitly debarring etymologies from consideration,²² and purely linguistic, untarnished with the "explanatory" pseudo-science of national-psychological stereotypes that was the conventional wisdom of the era. (Pokorny's work, by contrast, is pervaded with these anthropological notions; indeed, Pokorny criticizes Morris-Jones for

²¹ Cf. Thurneysen's comment: "[Celtic zero-copula sentences] cannot be compared with the nominal sentences of some Semitic languages" (1946:494).

²² "There are many Welsh and Irish words which cannot be explained from Aryan roots, and some of them may possibly be derived from a pre-Celtic tongue, but I have never pretended that I have discovered the origin of any one of them" (1904/05:278); cf. also 1900:638.

failing to take them into account (!), for carrying on his investigations "ohne die psychologischen sprachwissenschaftlichen Unterlagen, die allein seine These glaubhaft machen konnten, genügend herauszuarbeiten" [1926:110].)

Morris-Jones saw his own research (1900:638, 1904/05:279) as inspired by and contributing to the anthropological tradition referred to at the end of sec. 2.1. The historical explanation he offers is squarely in that tradition: the pre-Celtic substratum in Britain was again identified with the "Iberians", part of the same Mediterranean race as the Hamites and hence likely to be akin to them linguistically as well; the Celtic invaders of Britain decimated the indigenous male population, intermarried with the women, and fathered children who spoke the new language but with the old, pre-Celtic speech patterns they learned from their mothers (1900:617-18). It was, to be sure, a move of some sophistication to run this Celtic-Iberian link via Hamitic: "It is reasonable to suppose that the various divisions of the Mediterranean race originally spoke allied languages. ... The peculiarities of neo-Celtic syntax, then, are derived, not indeed directly from [the Hamitic languages], but from a lost language [Iberian?] allied to these" (1904/05:279). Another point betokens sophistication as well: the proposal that the largely analytical neo-Celtic syntactic prototype, found in Modern Welsh and Irish alike despite major phonological divergences, was actually present in the ancient spoken language. "It is the characteristic of the language of the people, and has been supposed to be modern only because it is not so apparent in the earlier literary language, which, besides being largely artificial, was based upon the dialect of a more or less Aryan aristocracy" (1900:641). This early

spoken language, of course, would not be preserved in the written tradition. Pokorny was to amplify these points at length (cf. sec. 2.3.1.2).

But the real originality of Morris-Jones's study inheres in the linguistics per se. As far as I can tell, no one before had laid out a concrete statement of the ineffable "genius" of Celtic vis-a-vis Indo-European; no one had carried out a detailed point-for-point comparison with other languages; and no one, apparently, had given any serious thought to Egyptian or Berber. Nor, apparently, had anyone appreciated the methodological importance of the sheer quantity of the resemblances. Morris-Jones, with a bachelor's degree from Oxford in mathematics (Parry 1972:23), was in a position to know the value of probability in arguments of this kind. As he said in a letter to the skeptical Kuno Meyer:

The adoption [in Goidelic and Brythonic] of the same form ... of analytical expression, not only in an isolated instance, but all along the line, in the two branches independently is so improbable as to be incredible. The mathematical probability against it is enormous. (quoted in Greene 1966a:126-27; italics in original)

Morris-Jones is here addressing intra-Celtic resemblances, not Celtic/Hamito-Semitic parallels; but the methodological point has general validity.

Morris-Jones's study pointed the question of a pre-Celtic substratum in a completely new direction, not only geographically but especially theoretically. What had heretofore been seen primarily as an anthropological issue could now be recast, for the first time, as a

problem in historical linguistic inference.

If Morris-Jones anticipated controversy, he was not disappointed. Scholarly disapproval ran the gamut from apoplectic splutter ("Welsh has neither part nor lot in Shem and Ham," Johnson 1904/05:160), to ad hominem attacks on M-J's scholarly qualifications ("Er ist im Irischen nicht zu Hause, weiss also nicht was keltisch ist; kennt keine indogermanische Sprachwissenschaft und kein Ägyptisch-Koptisch," Heinrich Zimmer [quoted posthumously in Meyer 1913a:107]),²³ to more serious responses.²⁴ The criticisms with substance to them are four:

a) The features M-J cites do occur in Indo-European, e.g. VSO (the Greek "schema pindaricum") and N-Adj word orders (Ridgeway 1907/08:50, cf. Johnson 1904/05:161-62). In his rebuttal (1904/05:280) to Johnson, M-J stresses the crucial difference between minority and dominant word orders: "[Johnson] might as well argue that there is no difference between a man and a dog because the dog may stand on its hind legs."

b) Ridgeway confidently asserts, as a "proved law", the principle that a conquered nation always imposes its language on the conquerers, "unless they come in great numbers and bring also with them wives of their own race" (Ridgeway 1907/08:45). Regardless of the validity of this "law", however, neither Ridgeway nor anyone else can do more than speculate on the sociological details of the Celtic penetration into

²³ Zimmer was not a man to mince words (cf. his comments on women's suffrage, 1910:1060), but the particularly brutal tone of this remark should be taken in context: the quote is from a patchwork book-draft found posthumously among Zimmer's "Nachlass", and as such might easily be a provisional, off-the-cuff formulation intended for revision.

²⁴ For an overview of these reactions see Pokorny 1926:109-11 and Greene 1966a:126ff., also various brief pieces in volume 1 of The Celtic Review (1904/05).

Britain.

c) Whenever foreigners have adopted an Indo-European language as their own, "the tense system is invariably broken up" in a sort of inevitable pidginization (Ridgeway cites "pigeon" English, 1907/08:47, 55). The survival in Celtic of the Indo-European tense system thus argues against M-J's substratal account (indeed, against any substratal account at all). --- Morris-Jones apparently made no response to Ridgeway, but one could be suggested. Ridgeway's attack is directed not against a substratal account per se, but against the kind of radical pidginization implicit in the social trauma of M-J's scenario (decimation of males, massive intermarriage, language upheaval in a single generation). However, substratal influence can proceed in many ways, and one could much more plausibly imagine generations of relatively trauma-free bilingualism, culminating in Celtic ascendancy but at the expense of slow syntactic contamination.

d) The most intelligent criticism, and by far the most sober in tone, came from the one man truly qualified to appreciate what Morris-Jones was talking about: Ludwig Christian Stern. Stern was in the unique position of being both a Celticist and an Egyptologist, having abruptly switched to Celtic philology in mid-career after two decades of work in Egyptology (including the production of a highly respected grammar of Coptic). In both fields he was acknowledged a master. Unfortunately, he wrote little on Celtic grammar (see Vendryes's 1911 necrology), and his only comment on Morris-Jones is restricted to a single page (Stern 1901:613).²⁵ Stern's point is typological: that M-J's parallels,

²⁵ In a letter of 1896, Kuno Meyer suggests to Morris-Jones that he solicit Stern's reaction to "your Hamito-Celtic article"; see S. Zimmer 1986:281-82. I do not know what, if anything, came of this suggestion.

however dramatic, might well recur in many other widely separated languages too, and are hence of limited probative value. (He goes on to suggest that such "analytical" features are the natural consequence of a language's loss or absence of inflection --- a less telling remark, given that neither Irish nor Egyptian [nor Semitic] is notably impoverished inflectionally, and that not all of M-J's features are particularly "analytical".)

When John Morris-Jones died in 1929 he was among the leading intellectual and cultural figures of Wales, and his passing was much mourned. Among the necrologies and later biographical studies summarizing his life's work, both by Welshmen and by continental scholars (cf. Vendryes 1929), there appears to be a curious omission: I have not come across a mention or hint of his early substratal studies. Yet by 1929 Pokorny's major article on the non-Indo-European substratum in Irish had appeared, reviving and praising Morris-Jones's pioneering work and lending it a new stamp of legitimacy from within the Indo-Europeanist camp. This silence bespeaks disapproval --- as if such substratal thinking, in the eyes of both the Welsh establishment and the mainstream of Celtic scholarship,²⁶ represented an embarrassment, a youthful indiscretion to be hushed up out of decent respect for the memory of the great man.²⁷

²⁶ Vendryes was far from an enthusiast of substratal explanation, as is clear from his 1937 lecture to the British Academy (pp. 345-46).

²⁷ There remains the remote possibility that Morris-Jones may himself have ultimately changed his mind about his own theory. My only indication of this is a parenthetical comment by D. S. Evans: "No one would now accept Sir John's theory in its entirety (he himself disclaimed it at a later date)" (1950:85a). This view appears to have been "oral lore" among Welsh scholars at Swansea around 1950, as Prof. Evans kindly informs me (p.c.). However, I have been unable to track down a single source, published or unpublished, written or oral, to corroborate this --- nor could Professor Evans himself, nor Huw Walters (M-J's bibliogra-

2.3 Julius Pokorny

If the Indo-Europeanist mainstream might dismiss Morris-Jones as an underinformed provincial (recall Zimmer's comment), the same could hardly be said with regard to Julius Pokorny, the co-creator and ultimate redactor of the great Indogermanisches etymologisches Wörterbuch. By 1926, the year of the first installment of his article "Das nicht-indogermanische Substrat im Irischen", Pokorny had been active in Indo-European and Celtic philology for almost twenty years, and Professor of Celtic at Berlin since 1920. His "substratophilia" was already a byword in 1917, when H. Gaidoz could acidly dub the pre-Celtic natives of Ireland "Homo Pokornius" (Gaidoz 1917:380, cf. Pokorny 1926:110). Pokorny's passion for substratal questions, to continue until his death in 1970, was sweeping in scope, embracing not only Celtic but also Illyrian and Indo-European as a whole. And the prehistory of Celtic itself involved a multiplicity of substrata: at various times Pokorny was to entertain simultaneous links with Hamito-Semitic, Basque, Caucasian, Finno-Ugric, and even Eskimo.²⁸

pher), nor Tomos Roberts (in charge of the M-J archives at Bangor), nor J. E. Caerwyn Williams (who knew M-J personally), all of whom graciously responded to my queries. And one piece of indirect evidence points in the opposite direction. In 1920, Julius Pokorny wrote to M-J: "My paper on the non-Aryan elements in Celtic is fast progressing; its first part will appear in the next number of the [ZCP]" (quoted in S. Zimmer 1986:284). This is the remark of one comrade-in-arms to another; it is hard to imagine Pokorny writing in this way had he known of a recantation. Arwyn Watkins (p.c.) informs me, on the basis of personal acquaintance with Pokorny in 1951, that the latter continued to retain the profoundest respect for Morris-Jones; in his articles (most recently 1964:76, praising M-J's "mit Unrecht totgeschwiegenen Arbeit"), Pokorny always refers admiringly to M-J's work, with no hint that Sir John ever had a change of heart. Could Morris-Jones have recanted without Pokorny knowing of it?

²⁸ For a bibliography of Pokorny's copious output see Schmeja's compilation in Meid 1967:323-32.

The idea of a Celtic/Hamito-Semitic connection was a leitmotif of Pokorny's long career. His massive first article (1926-29) covers the ground thoroughly in 150 pages spread out over five installments (the fifth ending very much in medias res, with a promised "wird fortgesetzt"). Successive treatments (1949, 1951, 1955, 1959, 1960a, 1962, 1964)²⁹ are largely refinements, incorporating more recent information, and/or recapitulations. Only Pokorny's treatment of the verb, somewhat neglected in his early article (as he himself states, 1962:132), was significantly expanded --- especially in response to work by such scholars as Bergin (1938a), Dillon (1943), and C. Watkins (1963) elucidating the initial position and the morphology of the Irish verb in strictly Indo-European terms. It is the 1959 article, which presents the main points of typological resemblance in catalogue fashion, that will be my point of departure in the data-oriented discussion of section 2.3.2.

2.3.1 Theoretical background

2.3.1.1 Substratal explanation

The basic theory behind substratal explanation as embodied in Pokorny's oeuvre is intuitively clear and attractive.³⁰ Speech

²⁹ On the verb: 1949, 1964; on impersonal idioms: 1951; on names: 1955; general surveys: 1960a (in English), 1962 (a German reworking of the same); concise linguistic summary: 1959. And on substrates underlying Indo-European: 1936, recapitulated in 1957. Also see 1953:132-34 for Celtic substratal discussion in the context of a bibliographical survey; and 1928b (summary of a conference report) for a two-page sketch of substratal issues covering both Celtic and Indo-European.

³⁰ For a comprehensive treatment of substrata in a modern perspective, see Thomason and Kaufman 1988 (discussed in sec. 7.2.2 below). Craddock (1969, Chapter 1) chronicles the fortunes of substratal thinking in European linguistics.

communities almost invariably are preceded on the territory they occupy by other speech communities, and these in turn by others; Ireland is no exception, as the archaeological record shows. Furthermore, it is both natural and common crosslinguistically for an older, indigenous language to exert structural and lexical influence on the new interloper; in Ireland we can see this happening in historical times with Hiberno-English. An attested language, then, is like an onion --- an accretion of layers which the substratalist strives to strip away one by one.

The primary methodological postulate of the method³¹ is that, when a language L1 has some feature of structure or vocabulary which resists ready explanation in genetic terms, then the puzzling feature is most plausibly to be ascribed to a substratum L2 --- rather than to coincidence, or drift, or an intricate elaboration in strictly language-internal terms. The case for substratal explanation is strongest when some demonstrably older language L2 having the desired feature is actually attested on the same territory as L1 --- the Cushitic substratum underlying Ethiopic Semitic, for example (Leslau 1945), or Caucasian vis-a-vis Armenian (Pokorny 1926:97). Far subtler are cases where no such indigenous substratum is available for inspection; its existence may be taken for granted,³² but its nature must be inferred. Here a second (and quite distinct) tenet of the method comes into play: that if

³¹ The distillation of Pokorny's substratal method presented here emerges from two sources: his own theoretical discussion, and plausible inferences based on his actual linguistic praxis.

³² Of course the existence of a substratum is not a logical necessity; one can imagine, perhaps, an isolated Pacific island where the present speech community is the only one ever to have existed. But such cases are of extreme rarity.

the given feature can be found in a geographically not-too-distant language L3, then the agreement should most plausibly be taken not as coincidental but as something historically meaningful, that is, as evidence that L3 or a cousin language represents one of the postulated substrata L2 underlying L1.³³ The case becomes stronger the closer the territory of L3 is to that of L1, and the greater the existing historical evidence for past links between speakers of L3 and the territory of L1. It may often happen that several different languages L3 are involved, each showing resemblance to L1 in some different feature(s). To the substratalist this is quite normal and unproblematical: each such L3 will correspond to a different posited substratum L2, a different layer of the onion. It is in this sense that Insular Celtic might be said to have simultaneous "affinities" to Hamito-Semitic, Basque, Caucasian, Finno-Ugric, and/or Eskimo.

The two methodological postulates presented above are of course simply assumptions, and historical linguists, depending on how attractive these assumptions appear to them, have tended to polarize into substratophiles and substratophobes, with mainline historical linguistics leaning toward substratophobia.³⁴ It will be instructive to compare the two approaches as regards both postulates, taking the present case of

³³ An extreme manifestation of this view can be seen, for example, in the work of Ernst Locker (1961, 1962), who sees traces of ancient "Sprachschichten" [strata] in individual typological parallels between individual languages of Europe and of sub-Saharan Africa --- English and Somali, for example, whose joint possession of a postposed asyndetic relative clause ("the man I saw") is "gewiss kein Zufall" (1961:148).

³⁴ Thomason and Kaufman discuss the long-standing bias against external explanation among orthodox historical linguists (1988:1ff., 57, 59), also mentioning the extremism of "substratomaniacs" (160, 352-53 note 8).

Celtic, Indo-European, and Hamito-Semitic as a model.

The substratophile accepts the first principle implicitly (even were there no "Hamitic-analogue" available to give the substratum a claim to concrete reality). His rationale is that substratal influence, as one of the most potent and best-documented forces underlying language change, is overwhelmingly likely to come into play whenever a new speech community overlays an old; given the sharp and diachronically unobvious divergence between Celtic and Proto-Indo-European, it would be astonishing if substratal influence had failed to play its usual dominant role. The substratalist sees his task as determining what that near-inevitable role must have been. The substratophobe, on the other hand, condemns this as uncontrollable speculation, refusing on methodological grounds to deal with unfalsifiable explanations no matter how attractive they may seem. To do so would be, in Martinet's words, "nothing but a confession of ignorance. It is ... preferable either to leave the question unanswered, or else to ask what structural factors in Early Celtic may have contributed to the development of the contrastive process" (1952:215; similarly C. Watkins 1962:1). In fact, though, linguistics abhors an explanatory vacuum; and so only seldom is the question actually left unanswered. Rather, the language-internal, structural route is explored in detail, frequently with insightful, if partial, results. These results are then typically taken, not as contributions toward an answer, but simply as the answer.³⁵ The substratophobe is sure that

³⁵ Watkins is completely explicit about this: "If, however, ... we can account for the peculiar development of the Celtic verbal system, as a direct and unmediated successor of the Indo-European verbal system, then the necessity for recourse to such hypothetical substrata simply disappears" (1962:1). Such a stance is typical of structuralism.

answering in such terms is superior to inventing a suitable substratum.

Yet this superiority rests only on methodological grounds, and has nothing whatever to do with fact. Could we but view the real historical development with an omniscient eye, the role of the unknown substratum/a might turn out to be far greater than the role of internal structural factors. Or it might not. A methodological preference for a certain mode of explanation is not a guarantee that that mode will be a more faithful exponent of the "truth" --- a dictum that holds equally for both substratophile and substratophobe. The problem at bottom is one of typological ignorance. Substratal action is undoubtedly very common crosslinguistically, but we have, I think, little secure knowledge about how common, or about how likely a language is to develop structurally deviant features in the absence of outside stimulus. To take a hypothetical, if we knew (on theoretical grounds) that substratal action was overwhelmingly likely to have played a major role in such cases --- as the extreme substratophile view would have it --- then a linguist's attempt to articulate the details of an unattested substratum would not be dismissed out of hand as "nothing but a confession of ignorance," but would instead find general acceptance as simply another form (albeit less reliable) of orthodox historical reconstruction. It is significant that the same broad question of external (contact) vs. internal (structural) causation continues to resurface perennially, as in debates in the 1970s over external (Lehmann) vs. internal (Antinucci, Vennemann) causes of word order change.³⁶

³⁶ See e.g. Lehmann 1978a, 1978b, Vennemann 1974, especially 1974:353.

Pokorny's own views here are unmistakable. Given the undisputed structural distinctiveness of Celtic vis-a-vis Indo-European, he has no hesitation in laying the burden of proof on those who reject a substratal account (1926:112-13; cf. also p. 95):

Wenn wir in gewissen Fällen im Zweifel sein könnten, ob einzelne Erscheinungen uralte idg. sind oder auf Einfluss eines primitiven nicht-indogermanischen Substrates beruhen, so werden wir meistens die letztere Möglichkeit für wahrscheinlicher halten müssen.

[If in particular cases we are in doubt as to whether individual phenomena are primordially Indo-European or stem rather from the influence of a primitive non-Indo-European substratum, we must in general take the latter option as the more probable.]

This extreme opinion, to be sure, reflects Pokorny's early view that Celtic was not a particularly archaic Indo-European language (1926:112). Later he was to modify this view (e.g. 1964:75), but his principled substratophilia never wavered. Between the lines there emerges a thinly-veiled impatience with other, language-internal approaches --- witness his cursory dismissal of Martinet's (1952) non-substratal account of consonant "lenition" in Western Romance as "far too complicated and quite unlikely" (1960a:238), with no further comment, or his repeated out-of-hand rejection of attempts to account for the Old Irish infixed-pronoun complex in Indo-European terms. And the possibility of independent convergence (i.e., coincidence) is backgrounded to the point of near-invisibility. This aversion to internal explanation makes sense in

light of the above discussion: it stems from a confidence (or overconfidence) that any such internal account, no matter how carefully argued and elaborated, is in actual historical fact unlikely to have played more than a minor part in a process whose fundamental impetus (so he feels) was surely external. And this feeling will be even stronger if the suggested internal account provides a "how" but no compelling "why".

It should be remarked in passing that orthodox diachronic linguistics, in condemning substratal explanation as "speculative", can itself hardly boast of methodological purity in this respect. The application of internal reconstruction as a tool for elucidating the pre-protostages of a language family has become an accepted and approved part of standard historical linguistic praxis, and represents the "cutting edge" of much of the best recent work on Indo-European. Yet such studies can be as much art as science: they are much less controlled and much more subjective than results based narrowly on the comparative method, and competing (and often mutually contradictory) accounts abound.³⁷ At the very least they typically embody an unprovable belief in diachronic entropy: that language change as a rule proceeds from greater to lesser regularity, and involves the disintegration or metanalysis of an originally coherent system whose (internal) reconstruction is the linguist's goal. Such a belief seems no less arbitrary than the dogma that substratal influence is always (or never) to be taken as the prime mover behind language change.

³⁷ Cf., for example, the variety of diachronic hypotheses regarding the Celtic verb, or the Hittite mi and hi conjugations, or the prefix vs. suffix conjugations of the Hamito-Semitic verb.

The gulf between the two approaches is just as wide with respect to the second principle --- that structural resemblances to nearby languages are very likely to have historical significance. Of course even the most ardent "a priori" substratophile would agree that the substratal argument is more persuasive when we do have some externally available linguistic reality L3 to which to anchor the posited substratum L2. Morris-Jones's Celtic-North African connection must therefore have been compelling to Pokorny, for Hamitic, could it be proven to have existed on British soil, would provide a substratal language of exactly the right kind. In far more detail than Morris-Jones, Pokorny's articles lay out point after point of resemblance between the language groups. Yet his writings show a curious opaqueness to the central crux of the whole matter: the matter of proof. As the substratophobe will be quick to point out, the mere demonstration of a commonality between two language groups does not ipso facto constitute proof of historical linkage; some cases are better than others. In this regard it is quite clear that the case for a Celtic-North African connection is different in kind, and an order of magnitude stronger, than the many other substratal links Pokorny advocated: the North African connection can appeal to some 17-odd shared features operating in block solidarity, while postulated links to Basque or Finno-Ugric involve far fewer structural affinities (see, e.g., 1960a:234-36, 238-40; also 1949:240ff.). As we have seen (sec. 2.2), the evidential value of this fact was not lost on Morris-Jones. But Pokorny, to judge from his writings, does not seem to have attached much importance to the point. For him substrates were not posited constructs in need of proof, but evident realities to be explored; an explanation in substratal terms was not so much a case to

be argued for as a plain fact to be displayed. The exposition of each individual point of resemblance might be accompanied by the litany "Es ist wohl kein Zufall," to confute the skeptic --- but the presentation as a whole is not.³⁸ It is hard to avoid the impression that Pokorny viewed all substrata with equal fondness. If the Celtic-North African connection stood out, it was because of its richer diachronic yield, not because it was more susceptible of proof.

2.3.1.2 Theoretical problems in substratal analysis

The bulk of European substratal work, from Schuchardt to Hubschmid, has tended to have a lexical slant, focusing on either toponyms or borrowed vocabulary. Indeed, Pokorny himself in the 1930s was to carry out a major toponymic analysis in his study of Illyrian. But the case for a Celtic/Hamito-Semitic link was entirely nonlexical. Numerous theoretical and methodological questions were involved here, and plainly some justification was in order.

Pokorny opens his 1926-29 paper (pp. 95-108) with an issue already noted above (under reactions to Morris-Jones): the question of which subdomains of language were felt to be susceptible to substratal influence and which were resistant. Here he voices a plea for flexibility. In cases of Sprachmischung [mixed languages], each instance must be evaluated in its own terms, within the context of its own individual historical circumstances (1926:96); absolute pronouncements about what

³⁸ This could, in fairness, be an accident of publication history: Pokorny's major 1926-29 article breaks off in the middle and has no concluding section, the obvious place to assess the argument as a totality.

must or cannot occur are ill-advised. Much will depend on how strongly the superstratal group holds the upper hand with respect to numbers, weaponry, social organization, and especially culture and literacy. When both groups are illiterate, the impact of substratal influence is sure to be great --- but not equally at every level of linguistic structure. As a general rule, Pokorny proposes that in such cases the dominant language will show near-complete ascendancy in lexicon and morphology, whereas in phonology and syntax ("innere Sprachform") the impact of the substratal language can be quite significant if the indigenous speech community has strong numerical superiority over the newcomers (97-98). His rationale is that lexicon and morphology are directly transferable ("ohne weiteres übertragbar sind", 98) --- are, in essence, learnable without significant interference from one's native language --- while phonology and syntax represent linguistic habits that run much deeper and are much harder to alter ("die sich nicht völlig austilgen lassen"). The direct corollary is that languages are unlikely to acquire vocabulary from a substratum.³⁹ The argument, as presented above, makes considerable sense in the abstract; but it founders against linguistic reality. Languages do borrow vocabulary from substrata: Ethiopic Semitic illustrates the point handily vis-a-vis Cushitic (Leslau 1945:79ff.). In the same vein Wilhelm Havers, in his discussion of syntactic borrowing (1931:133-44),⁴⁰ cites Rohlf's study (1930:38-

³⁹ Pokorny maintained this view throughout his career; thus in 1959 he writes: "Am wenigsten ergiebig ist natürlich der Wortschatz.... Ausserdem ist der Wortschatz dasjenige Element der Substratsprache, das am wenigsten einer Erhaltung günstig ist" (1959:161). For much the same point see Thomason and Kaufman 1988 (39, 133-34), who argue that the Ethiopic case (below) represents a mixture of substratal and borrowing influences.

⁴⁰ Including brief discussion of Pokorny (138-39).

42) of Greek substratal influence on the Italian dialects of southernmost Italy --- considerable in the sphere of syntax and lexicon, nonexistent⁴¹ in phonology (138). (Greek, to be sure, was originally very much a Kultursprache, with a written literature and a great deal of prestige, conditions Pokorny explicitly excludes in 1926, if not in 1959.) It seems evident that Pokorny is here attempting to theoretically justify, and even to predict, the non-appearance in Insular Celtic of any clearly Hamitic lexical stock. But the argument is not as compelling as he would like.

Another difficulty with substratal explanation of the Insular Celtic peculiarities is the fact that the degree of typological match with the posited Hamito-Semitic substratum, and of typological deviancy vis-a-vis Indo-European, is greater for later Irish than for Old Irish, increasing over time as the language gets further and further away from the prehistoric era when the putative substratal contact was a sociolinguistic reality.⁴² Such action-at-a-distance seems paradoxical, and was anathema to substratophobes like Jespersen, who categorically rejected the idea that substratal influence "may show itself several generations after the speech substitution took place" (1922:200). Pokorny's response (1926:102ff.) emphasizes the difference between the written record and spoken reality, and appeals (without of course using the term) to a kind of diglossia in ancient Ireland. The "high" Irish of the Celtic nobility, who spoke it as their inherited Indo-European

⁴¹ Or apparently nonexistent --- Rohlf's presents and rejects several possible instances (1930:41).

⁴² For further discussion of such chronological difficulties see sec. 7.4.

tongue, was surely not identical to the imperfectly learned Irish spoken by the non-Indo-European masses. But it is only the former, handed down orally over the centuries by a conservative bardic and legal tradition and later propagated in writing by a conservative scribal order, which has come down to us. It is precisely the latter, on the other hand --- if we had access to it at all --- which we would expect to show the strongest substratal influence (though of course attested Old Irish itself already shows considerable typological change vis-a-vis Continental Celtic). This suppressed "vulgar" layer of the language could only emerge into the documentary record when the continuity of the scribal tradition was broken --- as occurred, for example, at the time of the 9/10th-century Viking invasions, the event which marks the beginning of Middle Irish. Conventional wisdom has it that the language underwent extremely rapid change during this period. For Pokorny, on the other hand, it was not the language per se which changed so much as its cultural embedment, allowing the partial emergence of a heretofore-invisible (i.e. unwritten) stratum of popular Irish speech. The same process would have repeated itself over the course of the Anglo-Norman invasions and consolidation (12-16th centuries), with even more of the "folk" speech coming to the fore as Modern Irish emerged (1926:103; cf. also Greene 1966b:12). In this way the influence of the substratum, always present in the vulgar language, would in a completely natural way convey the appearance of growing stronger over the centuries.

The paradoxical "action-at-a-distance", accordingly, is just an illusion. As a methodological point this seems to me extremely telling; Havers concurs (1931:139). Pokorny, however, presents two other arguments which are much less successful. One appeals to the generations-

long spread of bilingualism in Ireland before the pre-Celtic language (whatever it was) ultimately became extinct. Over this period the influence of the substratal language would have increased progressively as more and more non-Celts imperfectly learned pre-Irish and imposed their own speech habits on it (as in such clear cases as Armenian vis-a-vis its Caucasian substratum). This argument makes sense, but only as long as the substratal language persists as a spoken reality. But the typological paradox concerning us here applies to a much later period --- not to prehistoric Celtic, but to the differences between the successive stages of attested written Celtic: Old Irish, Middle Irish, Modern Irish. Pokorny's present argument would thus have force only if these successive stages of written Celtic corresponded to stages of spoken Celtic in which the substratal language was still a living reality and Celtic/pre-Celtic bilingualism still quite ordinary. This seems highly unlikely (see e.g. Adams 1980:52, and sec. 3.1 below). --- Pokorny's final argument, a biological appeal to the genetic inheritance of linguistic tendencies ("die Vererbung bestimmter sprachlicher Tendenzen", p. 106), is explicitly recanted ten years later as "unbeweisbar und unwahrscheinlich" (1936:70).

As a matter of methodology, it would clearly be useful to have a way of inferring substratal action from observable grammatical behavior, and Pokorny indeed proposes such a diagnostic. A grammatical phenomenon often exhibits vacillation and resists clearcut formulation in rules --- and such cases, Pokorny suggests, are highly likely to result from Sprachmischung (1929:241). An analogous argument holds if the language as a whole wavers between two syntactic types. Thus in his 1936 article, discussing not Celtic but Indo-European, Pokorny observes that Proto-

Indo-European shows signs of both the "anreihend" and the inverse "unterordnend" type (see sec. 2.3.1.3 below); he concludes: "Derartige Diskrepanzen lassen sich nur durch Sprachmischung erklären" (1936:88 [= 1957:84]). --- The logic behind this, nowhere justified, is hard to fathom, except perhaps as yet another manifestation of Pokorny's extreme substratophilia. Of greatest interest from a modern perspective, however, is the striking similarity between this line of thinking and that embodied in Lehmann's theory of the dynamic of word order change: languages which are intermediate in type between "pure OV" and "pure VO" (i.e., "inconsistent") are ipso facto deemed to be in the process of change.⁴³ For both Pokorny and Lehmann, synchronic non-adherence to an ideal type is taken as a diagnostic of the course of a language's diachronic development --- though the particular conclusion to be deduced of course differs markedly in the two cases.⁴⁴

A recurrent theoretical problem in substratal analysis concerns the proper treatment of cases where a pattern in a language L1, putatively the legacy of a substratum L2, in fact can be shown to exist already in embryo in the parent language L1' of L1. Here any appeal to a substratum might seem intuitively unnecessary and methodologically wrongminded (Occam's Razor). This conclusion is specious. Occam's Razor can shave too close; it provides a mechanical argument against any proposed instance of multiple causation in diachronic change. In fact syntactic borrowing, proceeding as it does through bilingualism, typically does

⁴³ See Lehmann 1978a, 1978b (e.g. 1978a:34, 1978b:398ff.).

⁴⁴ In this connection recall also the remarks made above with regard to internal vs. external causes of language change, where again an insightful parallel can be drawn between Pokorny and Lehmann.

involve a structural resemblance and a sociolinguistic resonance between a majority pattern in the lending language L2 and a near-homologue in the receiving language (L1', then L1). The once-embryonic or minority pattern of L1', reinforced (through bilingualism) by pressure from L2, expands over time to become the dominant configuration of L1 --- a prototype case of multiple causation. This problem is an old one in linguistics, going back via Havers as far as Brugmann, who on at least two occasions (1904:45, 1917:54-55; see Havers 1931:134) proposes substantially the analysis just presented. Pokorny was himself well aware of the issue (1949:237, 1964:79) and of the explanatory tradition he was heir to, and his answer is in all essentials Brugmann's. Thus Brugmann (1904), quoted by Havers (1931) and in turn by Pokorny (1964):

Eine gewisse Übereinstimmung war schon von Anfang an da, aber was bei dem einen Volk nur okkasionell und nur in Ansätzen vorkam, war bei dem andern usuell und in grösserer Ausdehnung vorhanden; infolge des Sprachverkehrs wurde nunmehr das auf der einen Seite erst in Anfängen Vorhandene ... weiterentwickelt.

[From the very beginning there existed a certain congruence {between the two usages in the two languages}, but what occurred only occasionally and inchoately in the one speech community was for the other the norm, and more widely distributed; as a result of linguistic contact, then, what was originally present in the first language only in its initial stages ... underwent further expansion]

Pokorny mentions this argument specifically with regard to the initial

position of the verb (1964) and the behavior of infixed pronouns (1949), both phenomena with parallels in Indo-European, as his critics were quick to point out (see sec. 2.3.3). --- Interestingly, Pokorny's thinking in this matter seems to have evolved over time. Thus in 1926 (p. 131) he quotes the following remark of Baudiš's (1923:123, 125):

The Neoceltic changes are a natural development of the older inherited Indo-European material; the non-Aryan influence could come in only so far that it perhaps predisposed the speaker to choose certain possibilities of development ... yet there might have been some other reasons as well.

Baudiš's diachronic scenario and analysis match Brugmann's exactly; yet here Pokorny reacts negatively, disapproving of the reservations ("Vorbehalten") inherent in Baudiš's statement.

2.3.1.3 Typology, innere Sprachform, and national mentality

In his approach to linguistic typology, Pokorny belonged to a tradition going back to Wilhelm von Humboldt, and represented in Pokorny's own time by the works of such linguists as Franz Misteli, James Byrne, and F. N. Finck (see bibliography). For these typologists, the form of a language, in particular its "innere Sprachform", was self-evidently a mirror of the form of a people's Geist. Nations' collective mentalities could vary along several dimensions, notably "Reizbarkeit" ("excitability" or "stimulability", with overtones of distractability). The structure of their languages would of course follow suit: for the greater the people's Reizbarkeit, the shorter their attention span, and the less the

ability of speakers to retain multiple concepts in the mind long enough to be compacted into higher-order units and emerge as complex and hierarchically organized linguistic structures. Thus peoples with low Reizbarkeit (Turks, American Indians) would naturally have densely structured languages, with well-knit multi-morphemic words (or even unwieldy "Satzwörter")⁴⁵ and elaborate sentence-level subordination ("Unterordnung"). By contrast, peoples with high Reizbarkeit would have shorter words, or words composed of morphemes only loosely strung together ("anreihend"), and little complexity at the sentence level --- a stream of linguistic bits running along in parallel with the conceptual bits flitting in and out of the excitable mind. Other typological differences, such as word order, also seemed to fit naturally into this framework. N-Adj order, for example, was viewed as loose appositional restatement, a succession of fleeting non-coterminous snapshots of the same object under different aspects ("the-house, the-big-[thing], the-white-[thing]") --- the natural manifestation of a high-Reizbarkeit national mentality. Adj-N order, by contrast, naturally went with low Reizbarkeit, for speakers had to be able to retain the adjectival concept in mind until the noun came along, and then put the two together --- the Adj could not properly be processed in the absence of the Noun (consider the manifold semantic possibilities of "good" in "the good X", an ambiguity which cannot be resolved until the appearance of the noun; cf. Finck 1899:65-66).

⁴⁵ On the other hand, both Pokorny and Wagner present "Satzwörter" as a characteristic of "anreihend" languages (see secs. 2.3.2 [3], 2.4.3 [21]). Humboldtian typological constructs can show considerable mutability from one linguist to the next; see sec. 2.4.1 below on the checkered career of "anreihend".

To European scholarship in the imperial heyday of the turn of the century, it was a commonplace that Africans and Celts were notably high in Reizbarkeit. Pokorny (1926:129-30) cites Byrne's authoritative pronouncements:

The Celtic race is distinguished among the Indo-European races by quickness of thought; and accordingly their language shows a tendency to break thought into smaller parts than any of the Indo-European languages. (Byrne 1885 II:182)

--- and regarding Africa:

Of all mankind, the genuine African races have the most quickness of excitability... [and correspondingly:] There are two characteristics which belong to all the purely African languages, a tendency to break speech into small fragments, and a readiness of the parts into which it is analysed to enter into combination with each other. (I:87)

It was only natural, then, that Celtic and African languages ought properly to be similar in structure.⁴⁶

Not merely the theory but the praxis of typology was affected by such views. Typological classification drew on two types of data.

⁴⁶ Pokorny's fondness for this vein of typological thinking apparently never waned. To be sure, appeals to national mentality and its correlation with linguistic type no longer pervade the analysis after the 1926-29 article; but in 1959 he still could refer en bloc to African "Negersprachen" (154), appeal to the "anreihend" character of Insular Celtic (153, 159), and seek an explanation for impersonal constructions in northern European languages in Finck's dictum that "der Mensch im eisigen Norden ein hilfloser Spielball der Natur ist" (164).

Intertwined with purely linguistic facts, and (apparently) coequal to them in importance and evidential value, were "insights" gleaned from what might be called a pseudo-science of national stereotypes. Very broad and amorphous constructs arose which, suitably interpreted, could be seen as naturally reflecting both the mental/cultural and the linguistic sides of the coin --- such macro-typological notions as a language's overall "subjectivity" (sense of subjecthood), or "passiveness", or "abstractness", or "fragmentariness". These concepts, oriented far more toward innere Sprachform than toward any concrete morphosyntactic features, were the fundamental forces driving and shaping 19th-century typological thinking, the counterparts to such modern-day macro-notions as "word order type".

A telling result of this macro-thinking was that languages might be judged similar in type even in the absence of precise structural parallels (in the modern sense), if they only could be seen as having features manifesting a common macro-type. It was in this sense that all African languages could be grouped together typologically, as relatively fragmentary languages with rather low subjectivity and abstractness (even Bantu, with effort, was subsumed under this rubric). Of course exact structural parallels were valued, and eagerly seized upon where they did occur. But their absence was no barrier to grouping languages together.

Pokorny's typological work suffers not infrequently from this syndrome: disparate phenomena will be coupled together in an attempt to demonstrate the structural similarity of two language groups. Two such apples-and-oranges examples, both from very late in Pokorny's career

(1959 and 1960a), will show how deeply ingrained this habit of thinking was in his approach to typology. The first (1959:158-59) concerns the simultaneously active and passive value of the Celtic verbal noun (à la "the shooting of the hunters"); as a "helpful" parallel, Pokorny adduces the fact that in Berber, one and the same (finite) verb form is very often both transitive/active and intransitive/passive (like English "break", "open", etc.). Such a comparison, which in today's typological discourse invites dismissal as a non sequitur, makes sense only when laid upon a presupposed macro-dimension of whole-language "passiveness". Secondly, in discussing the piling up of preverbs in the Irish verb (1960a:237, 1962:133), Pokorny adduces as a parallel the phenomenon we now call serial verbs, common in West African ("Sudanic") languages and found in some "Hamitic" as well (Meinhof 1912:28), whereby semantically complex verbs are broken down into a chain of simpler verbs ("bring" = "take-come", etc.). The Irish word for "servant" features multiple preverbs:

timthirithid < *to-imbi-di-ret-yatis,

glossed by Pokorny as "he who comes up (to-) to his master, goes around (imb-) him and comes back (di-) again". The action is thus decomposed psychologically into bits, just as in serial verb constructions. Again the parallel makes sense only on condition that the notion "tendency to fragment speech into small conceptual bits" is to count as an important typological macro-parameter.

Typologists today, in reading Pokorny, have to expect to keep coming up against this very alien theoretical worldview --- to recognize it for what it is, to use it in understanding Pokorny's argumentation, and to abstract away from it in appropriating his data for our own use.

Such an "abstracting away" could potentially be quite tricky: with some scholars,⁴⁷ theory pervades data so thoroughly that one cannot abstract away from the theory without the fabric threatening to unravel completely. Fortunately, in the present instance the difficulty is minimal. Examples like the two given in the previous paragraph are extreme cases; Pokorny's comparanda are usually much more tightly constrained, and can be easily lifted out of their theoretical embedment.

Pokorny's typological worldview, in positing a principled, natural-seeming linkage of linguistic features and ethnological features (culture, national mentality), has another and quite peculiar consequence --- one which has not, to my knowledge, been noticed before. In effect, his metatheory undercuts the very point he is trying to demonstrate! The issue is methodological, and again concerns the provability of the posited Celtic-North African connection --- the heart of the matter for us, if not for Pokorny. The best possible case for such a connection would be one which rested on observed affinities both in linguistic factors and in non-linguistic features --- archaeological, anthropological, ethnological (including, for Pokorny, considerations of national mentalities). The value of this dual basis inheres precisely in the fact that two independent arguments point to the same conclusion. But if language type and national-mentality type are felt to be natural correlates of one another, then the dual nature of the confirmation is weakened, and the whole proof along with it. Paradoxically, the more coherently and naturally the various factors (linguistic and nonlinguistic) mesh in a harmonious ensemble, the less their evidential value.

⁴⁷ E.g. Heinrich Wagner, as we will see below.

The principle is a familiar one: quirky, unnatural resemblances are the best proof of historical links between languages, and the more the better. This dictum is standard in genetic diachrony, but applies equally to cases of contact influence or borrowing, and indeed wherever it must be demonstrated that a resemblance between two languages (or language groups) is not a case of independent convergence but draws upon a shared history.

2.3.1.4 Africa and Hamitic

Pokorny's image of the genetic and classificatory map of Africa also reveals him as a man of his times. His views draw on those of the great Africanist Carl Meinhof (1912), according to whom Hamitic languages could be found over large parts of the continent (extending to Maasai and even Hottentot). Originally the speech of the North African "white" race, the Hamitic languages had over the centuries penetrated throughout Africa, often mingling with the "Negersprachen" of Black Africa to produce Hamitic-like Mischsprachen. This picture, which simplistically projected typological and putative racial isoglosses onto genetic diachrony, has been debunked by Greenberg (1966b); but it was the accepted conventional wisdom through the first half of this century. On top of this Pokorny, as indicated above, believed that all African languages were structurally similar in virtue of their "anreihend" typology (1926:129). The result was that his Celtic-Hamitic comparanda were not limited to Mediterranean languages (Semitic, Egyptian, Berber), or even to "narrow" Hamitic, but would range all over Africa, encompassing Bantu or "Sudanic" (West African) languages at will. Pokorny

explicitly acknowledges and justifies this procedure (1926:135-38), on several grounds: that Hamitic and Sudanic languages themselves shared a common substratum; that Hamitic speakers in their pan-African expansion had left behind deposits of Hamitic features even in non-Hamitic languages, linguistic fossils that could provide otherwise unrecoverable insights into Hamitic protohistory; and finally, quite apart from any considerations of genetic relatedness, that the "primitive" languages of Black Africa might by their very primitiveness shed light on the primitive Urhamitisch language type(!).

Half a century later these views seem by turns funny, fantastical, and/or offensive. Fortunately, they are almost never crucial to the actual linguistic substance of Pokorny's argumentation, which only seldom appeals to comparisons with sub-Saharan Africa. As with Pokorny's typological views, here again we must separate out and abstract away from various strata in his thinking in order to recover what there is of value. And here too we can learn about the danger of uncontrolled substratophilia from the nature of Pokorny's argumentation. Diachronic linguistics for him unfolded against the dramatic backdrop of vast population movements, stratum upon stratum, language overlaying language --- a prehistory which, in Pokorny's writing, tends to be reified into a definite shape beyond anything justifiable on the basis of the mutable and skimpy archaeological and anthropological evidence. It was as if reconstructed prehistory were real the same way history was real, and could explain linguistic change with the same validity as "true" history. And, of course, if the reigning view of African prehistory could be relied on in this way, then Hamitic elements and/or the anreihend macro-type were sure to crop up wherever one turned in Africa. It was

just a question of ferreting out their many manifestations, all of which would be legitimate comparanda vis-a-vis Celtic.

2.3.2 Points of similarity

When all is said and done, we see that most of Pokorny's theorizing --- despite his chastisement of Morris-Jones for the "unsprachwissenschaftlich" character of his nontheoretical approach --- today represents only so much deadwood to be stripped away. The point about the "delayed" effects of a substratum, ultimately emerging when a previously suppressed "folk speech" comes into its own, constitutes an original and insightful contribution; the point about nontransference of vocabulary in a substratal context also has a measure of validity. The rest yields much insight into the history of linguistic thought but almost none into real linguistic explanation. We are left, as with Morris-Jones, with the empirical presentation of typological facts --- and here, fortunately, a wealth of truly valuable material emerges.

The presentation in this section follows Pokorny's 1959 article point for point, with relevant cross-references to the other articles (especially that of 1926-29). I do this not only for expository convenience, but also to demonstrate in capsule form how the linguistic side of the problem, as a totality, appeared to Pokorny at a late, mature stage in his career. My presentation has a twofold aim: not only to lay out a factual summary, but also --- and sometimes principally --- to flesh out the above discussion of Pokorny's method with examples and criticism.

1) The Indo-European phonological system has been restructured more drastically in Celtic than in any other branch except Armenian, what with lenition and (in Irish) palatal and rounded coloring of consonants (cf. 1960a:238; also Zimmer in Meyer 1913a:107-109). Pokorny draws a parallel to palatal and rounded quality in Berber, where phenomena analogous to Irish fiur < *wirū can occur: "Es ist bestimmt kein Zufall, dass auch im Berberischen sogenannte i-haltige und u-haltige Konsonanten eine grosse Rolle spielen" (1959:153). But Berber has such "coloring" only with gutturals (k-kw-ky, etc.), as is quite clear from the source Pokorny cites (Stumme 1899:10); in this respect Berber does not look greatly different from Indo-European itself. Further, the Berber phenomena cited do not (apparently) involve a consonant absorbing a proto-vowel and being colored by it, but only a low-level phonetic metathesis (kw/wk). The Irish colorings, by contrast, pervade the entire consonantal inventory, and explicitly arose from absorption of earlier vowels. Pokorny has taken a minor partial resemblance and blown it up out of all proportion --- an all-too-characteristic move.

2) Irish shows strong "anreihend" tendencies that set it apart from Indo-European; cf. 1926:127-144. In his 1959 article Pokorny cites only two such features (vi, vii below), but in 1926 a profusion of phenomena appear under this header, each rationalized in a different way as manifesting the alleged "anreihend" or "loosely linking" macro-type (sec. 2.3.1.3) --- such features as:

i) The heavy use of preverbs, semantically fragmentary morphs which are an obligatory part of the verb yet are not fused into it (Finck 1899:41-42). (For Finck, this is reminiscent of the Dantu noun class prefixes, which are likewise semantically minimal yet indispensable

elements of word formation.) The Irish usage is asserted to occur "in weit ausgedehnterem Masse, als in den anderen idg. Sprachen" (130). No mention is made of the similar use of separable preverbs in un-anreihend German.

ii) A new-information-first principle (verb-initial order) and much prolepsis and afterthought, all a reflection of a fragmentary mode of expression (per Baudiš 1913a:311-12).⁴⁸

iii) (per Baudiš:) The order Noun-Attribute (see sec. 2.3.1.3, and (4) below).

iv) (per Baudiš:) Non-existence of -nt- ("active") participles as an inflectional category of the verb, a major deviation from the Indo-European type (cf. Meillet 1964:373). This argues (so the claim ran) that Celtic lacked the means for subordination, an earmark of the anreihend type; presumably subordination via verbal nouns, extremely common in Celtic, did not count. No mention is made here of the IE -to- ("passive") participles. In most Celtic languages these have become purely adjectival, but in Breton they survive in their inherited function as participles --- an indication that early Insular Celtic could realize subordination through participial means. See also (7) below.

v) A tendency, in conjoined phrases in Modern Irish, to repeat a shared

⁴⁸ The Czech linguist Baudiš was himself deeply interested in the deviant character of Celtic vis-a-vis Indo-European, and dealt with the subject in several articles (e.g. 1913a, 1913b, 1922, 1923, 1926). Though definitely sympathetic to substratal explanation as a contributing factor (1922:33, 1923:120ff., 1926:216-17), Baudiš cautioned against overzealous substratophilia, and explicitly refrained from identifying the pre-Celtic substratum with any other linguistic group (1926:216). Rather, his focus was on working out the concrete details of a plausible path from Indo-European to Celtic, casting the development more broadly within the macrotypological categories of the era (cf. sec. 2.3.1.3): Celtic was a "fragmentary" language (1913b:380-81); it exploited the "emphatic" side of Indo-European (1926:216).

article, possessive pronoun, or preposition (as if "with a knife and with a fork" instead of "with a knife and fork"). No mention is made of the fact that the opposite tendency is also in evidence in Modern Irish: a sequence of conjoined adverbs may be marked only once for adverbiality (with go), the mark appearing on the first conjunct (go Adj Adj).

vi) The common use of the paratactic syntagm "and [Nominal Clause]" in place of the subordination normally found in IE (the type "Did you think I would drink this, and my wife dead?"). Here a telling parallel exists in Arabic, and (allegedly) in Egyptian. However, Pokorny fails to note that in Arabic the paratactic "and"-clause can be either nominal or verbal. And his proposed translation (1926:139) of the relevant Egyptian particle(s) ist, ti, isk as "and" finds no support in Gardiner's grammar (1957:68-69, 177, 182).

vii) A strong tendency in prose toward short paratactic sentences with little subordination. In 1926, Pokorny illustrated this with four pages (132-35) of text examples drawn from Old and Modern Irish, with parallels in Egyptian and Berber; the point is repeated in 1959, with Hebrew and Arabic added in. --- But there is a curious kind of ten-year amnesia at work here. In the interim between 1926 and 1959, Pokorny (1949:235) brings up Thurneysen's criticism that many of Pokorny's allegedly linguistic features are actually far more an expression of cultural and literary norms. Thurneysen points out that Irish literature, unlike most other European literatures, did not begin as a translation vehicle for classical originals. Hence

manches von dem, was Pokorny als vom indogermanischen Brauch
abweichend empfindet und auf hamitische Grundlage zurückführt,
darauf beruht, dass der irischen Literatursprache nicht die

bei den Griechen und Römern ausgebildete zugrunde liegt.

(Thurneysen 1930:428)⁴⁹

[much of what Pokorny perceives as deviating from IE usage, and traces back to Hamitic, results from the circumstance that the literary standard which developed among the Greeks and Romans does not underlie literary Irish.]

Not only does Pokorny cite this quote without rebuttal (in an article otherwise full of polemic), but he even amplifies it by volunteering illustrations of what Thurneysen must have meant --- points not specifically detailed by Thurneysen himself --- and the items thus adduced include (inter alia) the domination of parataxis over hypotaxis in Irish, and the preference for short sentences! Though Pokorny does not say so, it is difficult to escape the impression that he is here implicitly accepting Thurneysen's view that these features are not so much linguistic as literary/stylistic traits. Not a hint of this, however, a decade later in 1959.

3) Insular Celtic shows a marked decrease in the autonomy of the word, a major deviation from Indo-European (cf. Meillet 1964:356) --- in Celtic it is groups of words rather than individual words which tend to act as units (cf. 1926:231-239). The manifestations of this tendency toward the "Satzwort" assume a bewildering variety, of which only two (i, iii) are mentioned in the 1959 article:

i) The fact of the initial mutations, a phenomenon which Pokorny is at pains to distance typologically as much as possible from Sanskrit sandhi

⁴⁹ Pokorny (1949) misquotes slightly.

phenomena; somehow mutation but not sandhi is felt to be a legitimate symptom of strong word-group tendencies. Pokorny calls the Celtic inflection-by-mutation "Gruppenflexion", and opposes it to IE "Stammflexion".

ii) Various instances in Celtic where several words are pronounced as a single group (1926:234-36), with a variety of alleged African parallels --- notably the Hamito-Semitic status constructus formation, and the Hebrew stop/spirant allophony, which (as in Celtic) can affect not only word-internal but also word-initial consonants occurring within a word group.

iii) The fact that questions tend to be answered, not with a word (here; me; yes), but in full sentences that usually echo the predicate of the question. Pokorny points out that this tendency is stronger in Modern than in Old Irish (1926:237; recall sec. 2.3.1.2). Here the African parallels draw on Egyptian (personal communication from Sethe), on sundry "Negersprachen" (Kpelle, Nubian; 1926:238), and (in 1959) on dialogues taken from Hanoteau's 1896 grammar of Tamashek (Berber). Little rationale is given for why this should count as a manifestation of a tendency toward "Satzwörter".

iv) Infixes pronouns and compound verbs (1926:239); see (12) below.

4) Regular VSO, N-Adj, N-Gen order, not found as a dominant order in any other old IE language but the norm in Hamito-Semitic.

5) A drastic weakening of the strongly "subjective" verb of Indo-European, where a clearly profiled concept of subject was an integral part of the verb. Thus impersonal constructions are a salient aspect of Celtic grammar, much more so than in other IE languages --- e.g.

is di Ultaib dom "I am an Ulsterman"
 (it)-is of Ulstermen to-me

or

do-rinne salann dona clochaib "The stones turned into salt"
 (it)-made salt of-the stones

(See also 1949:240-42, 1960a:234-35, and Vendryes 1956.) Noteworthy, too, is the conversion of the IE passive to an impersonal in Celtic. Pokorny seems to waver regarding the substratal origin of these impersonal usages. He finds parallels in Basque and Caucasian (ergativity; cf. 1960a:239-40), both (per Pokorny) belonging to an eastern Mediterranean linguistic stock brought to Britain by the Beaker-Folk migration --- and also in Eskimo (1955:63) and in Norse and Finnish (1960a:235), the last three perhaps reflecting traces of an "Upper Palaeolithic" circum-polar substratum. Whatever the case, attribution to Hamito-Semitic is not seriously at issue; the few Semitic parallels adduced (1959:155) are not persuasive. Note that the variety of crosslinguistic phenomena united here under the rubric "impersonal" are directly comparable only in their capacity as manifestations of the assumed macro-feature "low subjectivity" (see sec. 2.3.1.3).

6) A major change in the tense-aspect system, with periphrastic tenses heavily featured --- and notably constructions of the type "he is at coming". Comparisons are made to Egyptian and Hausa.

7) Lack of the active ("present") participle as a syntactic category; its function is filled with verbal-noun constructions (again, "he is at coming"), as in Berber and Egyptian. No mention is made of the fact

that Semitic and Egyptian also have active participles (the so-called Berber "participle" is a misnomer, being restricted to relative clauses).

8) Irish, unlike all the other old Indo-European languages, has two verbs "to be", a copula and a substantival verb.⁵⁰ Parallels can be found in Hamitic, Finno-Ugric, and Spanish (the last going back to "dasselbe vorkeltische Substrat").

9) Insular Celtic deviates from the IE tendency toward "abstractness" --- another presumed typological macro-dimension (cf. 1926:239-59). Under this header are such features as the following (only i, ii are mentioned in the 1959 article):

i) Constructions of the sort "there is X upon-me", etc., instead of "I have/feel X" --- notably "There-is a-shilling by-me upon-you" = "You owe me a shilling" (cf. Pokorny 1951, with parallels in Semitic and Egyptian). The alleged concreteness is apparently deemed to stem from the presence of the spatial preposition. Curiously, this phenomenon was not mentioned in 1926 under this header.

ii) An uncharacteristically heavy use of possessive pronouns, especially with body parts and clothing; "I put my hand in my pocket" (like English), not "I put the hand in the pocket" (the asserted IE norm, as in German). Pokorny (1926:251-52, quoting Sayce) links this with the phenomenon of inherently possessed nouns found in "primitive" languages, an expression of "the primitive inability to separate the particular

⁵⁰ The point is not etymological but functional --- not merely the copresence of reflexes of IE *es-, *bhew(H)- but their repartition into two competing paradigms, each expressing a distinct set of nuances of "to be".

from the universal." The English usage is laid to a Celtic (ultimately pre-Celtic) substratum (1926:253); the fact that the same thing occurs in Dutch prompts Pokorny to posit substratal influence already in Gaulish (1929:248) (!).⁵¹ Parallels exist in Egyptian (per Kurt Sethe).

iii) The use of the concrete term ceann "head" in counting (like numeral classifiers); likewise the use of cuid "portion" in apparently redundant contexts, such as (1926:239-40, 243):

a cuid éadaigh	"her clothing"
a gcuid capall	"their horses"

This "unnecessary" concreteness shows the Irish deviation from IE abstractness. Here Pokorny's parallel is to Berber and to the (Niger-Congo) African language Gola.

iv) A variety of counting usages, such as "man of the men" in the meaning "one of the men, one man, a man"; the concrete noun ("man") is repeated instead of an abstract pronoun (1926:241, 246). --- In this context Pokorny notes that "wir würden ... viel eher einfach den unbestimmten Artikel gebrauchen" (247); in the next paragraph he appeals to "unserem Sprachgefühl". This explicit appeal to modern European norms (one is tempted to say "modern German norms") is all too common; see also the discussion of "my hand in my pocket" above. It is never made clear why modern usage or Sprachgefühl should be relevant to ancient IE syntax, a particularly confusing point in the present case, as Pokorny

⁵¹ This ad-hoc appeal to pre-Celtic substratal influence even on the continent --- something at variance with the specialness of Insular Celtic --- shows the substratalist's enthusiasm at work. Pokorny suspends one of the central tenets of the exposition in order to preserve a substratal explanation of this one small item. Perhaps coincidentally, it is precisely at this uncomfortable point that Pokorny's five-part article breaks off.

was of course aware that no old Indo-European language had an indefinite article. Is he implying that the modern languages of Western Europe somehow have preserved the inner essence of ancient Indo-European especially well?

10) Conjugated prepositions, with good parallels throughout Hamito-Semitic, also in Finno-Ugric (cf. 1928a:385ff.).

11) Pokorny remarks, as had Morris-Jones, the precise parallels between the uses of the Preposition "in" in Welsh and Egyptian --- as locative preposition, predicative particle, and constitutive element in periphrastic verbal tenses. The Irish type "he is in-his doctor" ("he is a doctor"), usually considered a late development (see e.g. Dillon 1927-28:324-26), "könnte ... alt sein" (1959:157).

12) Like Morris-Jones, Pokorny noticed the close structural parallelism between Celtic and Berber in the placement of clitic object pronouns: suffixal when attached to a "bare" verb, but infixal when the verb is accompanied by a preverb or conjunct particle; schematically,

V-Obj vs. Prev-Obj-V

(1928a:382ff., 1949:235ff., 1964:76ff.). To those who propose to explain this patterning through appeal to Wackernagel's Law, as a purely IE-internal development --- Vendryes and Dillon (see Pokorny 1949), Calvert Watkins (see Pokorny 1964) --- Pokorny's main response is "why?" Specifically:

i) Why did Celtic alone of the IE dialects fail to develop independent object pronouns (or at least not until a much later period)?

ii) In IE, Wackernagel's Law constrains clitic pronouns to second position, but allows a variety of elements to appear in first and third

positions. Why only in Celtic was this flexibility lost, so that the only patterns commonly allowed (ignoring the rare instances of tmesis) were the unverbated sequences Prev-Obj-V (and V-Obj, when no Prev exists)?⁵²

--- Substratal influence, on the other hand, can explain why this particular pattern, of only minor importance in Indo-European, was singled out to become the Insular Celtic norm.

Alongside this purely structural comparison, however, Pokorny also proposed a cluster of Berber-Celtic etymological equations involving infixed pronominal elements, a notion he pursued over the decades (still in 1959) with the tenacity of an idée fixe. The etymological link, as first advanced in 1928 (1928a:384), is a tentative suggestion involving the morph d; by 1949 all hesitation has vanished,⁵³ and the equation has expanded to encompass two more particles (found in both Basque and Berber); in 1955 (p. 63) Pokorny angrily dismisses the pointed criticisms of the Basque scholar Bouda (Bouda 1949); but in 1959 the comparison again involves only d. The details, as laid out in Pokorny 1949:242-45, are as follows:

i) Berber has a number of (near-)homonymous particles d, possibly related etymologically (as argued by Marcy 1939), with clearly defined functions: a preposition "with", a predicative particle, a tense/aspect

⁵² A third explanation is also put forward: the alleged unreasonable-ness of an account whereby the most weakly stressed element dictates terms, as it were, to the rest of the sentence (1949:237; in Pokorny's picturesque phrasing, "Das Wackernagelsche Gesetz wirkt doch nicht wie ein chemisches Gesetz").

⁵³ "Ich zweifle heute nicht mehr daran, dass auch die lautliche Übereinstimmung [von d] mit dem Irischen auf keinem Zufall beruht" (1949:236).

particle, a relative particle (in some dialects only), and a proximal deictic particle "here/hither". The last of these is part of the syntagmatic complex of elements which moves as a block to infixed position in the presence of a preverb (notably with the Neg and Tense/Aspect preverbs, and in relative constructions). --- In Irish, there is a recurrent submorphemic element d of opaque synchronic function and uncertain etymology (< IE *idhe (?), cf. Lewis & Pedersen 1961:243, Thurneysen 1946:325). Under well-defined conditions it occurs suffixed to the negative (na-d) and to various conjunctions (con-id "so that", ma-d "if", cia-d "although"), and also appears as a meaningless formative element in the "series C" set of infixed pronouns, a series which is semi-specialized for use in relative clauses. --- The functional resemblances between the homophonous Irish and Berber particles are beguiling: the notions "negative", "relative", and "infixing" play a role in both cases, yet without any clear or systemic parallelism emerging.

ii) Old Irish forms relative clauses in a variety of ways, one involving lenition of the verb stem, the lenition being the assumed reflex of an earlier vocalic particle (Thurneysen 1946:323). In corresponding function Welsh has the relative particle a, a form which Pokorny posits as the pre-Irish original as well. The particle is difficult to etymologize in Indo-European terms; where the IE relative *yo does show up in Celtic, its reflexes (quite reasonably) do not assume the form a. --- Some Berber dialects also have a relative particle a (elsewhere ay); Pokorny further suggests that Basque does as well (see Bouda's response, 1949).

iii) Another Old Irish relative-clause type (used for oblique RCIs) involves nasalization of the verb stem, the assumed reflex of an earlier

nasal particle. This proto-particle, too, is etymologically vexed (Thurneysen 1946:323-24, Lewis & Pedersen 1961:243). --- In Berber, a sub-morphemic element n appears in subject-relative clauses, normally suffixed to the verb (thus forming the so-called "participle"), but in infixed position when the Neg preverb occurs:

Verb-n vs. Neg-n-Verb "(he) who (NEG) verbs"

This participial n is not one of the standard "infixing" elements, for in many dialects it assumes an infixed position only with Neg, retaining its suffixal status with all other preverbs (see Basset 1949:35). It may be etymologically linked to a homophonous distal deictic particle (thus Marcy 1939, cf. also Basset 1949:35), which is a standard infixing element like its proximal counterpart d (above). Pokorny also adduces a Basque morph en, allegedly functioning as both genitive and relative-clause marker, which he sees as etymologically linked to the Berber morpheme (again see Bouda 1949 in rebuttal).⁵⁴

With such a gossamer tissue of partial similarities, it will inevitably be a matter of taste whether one sees the beauty of the web or the thinness of the strands. For the substratalist, there is too much to be passed over as coincidental: a Hamitic element is manifestly revealed as present in both Basque and Insular Celtic. Pokorny ties things together neatly (1949:245):

⁵⁴ In fact, there is an important morphological difference between genitive -en and relative -(e)n in Basque: the e is obligatory in the former, variable (under morphophonological control) in the latter (Rebuschi 1984:129). Bouda puts it differently, taking the genitive ending to be -e and analyzing -n as a general-purpose embedding marker. On either view, Pokorny's equation of the two morphemes is seen to be an oversimplification.

Im Keltischen und Berberischen, zum Teil auch im Baskischen, finden sich nebeneinander die drei Relativzeichen a, d, n, die nicht nur in der Form, sondern auch in der Funktion übereinstimmen, was kaum ein Zufall sein wird.

[In Celtic and Berber, and in part also in Basque, there occur alongside one another the three relative markers a, d, n, agreeing not only in form but also in function, which can hardly be a coincidence.]

But his argument, taken as a piece of rhetorical exposition, is structured as a loose juxtaposition ("Anreihen"?) of partial parallelisms, an argument through pregnant hints --- an expository strategy which is, unfortunately, not at all unusual in Pokorny's substratal writings. And the case, of course, is as elusive as it is seductive. Pokorny himself admits that "so kurze Worte nur mit grösster Vorsicht verglichen werden dürfen" (1949:244), then proceeds to ignore his own dictum; indeed, a, d, n are among the commonest sounds cross-linguistically. The etymological link among the various Berber d-particles is (as best I can tell) not solidly established; the Berber a-relative would appear to be an alloform of ay, which would founder on the same phonological difficulty (the /y/) as does IE *yo if posited as the etymon underlying Insular Celtic *a; as for Basque, Bouda's critique seems to the point. Functionally, the non-parallels vie with the parallels: Berber n occurs only in subject-relative clauses, Irish *n only in non-subject relatives; Berber n is usually suffixed, Irish *n always infixes; the functional clarity of Berber d (taken in its several distinct uses) contrasts with the functional opacity of Irish d in all its uses, precluding any real functional comparison. The form/function similarities, however

tantalizing, do not meet normal standards for an etymological equation.

13) Unlike Indo-European but in close agreement with Hamito-Semitic, Insular Celtic has no relative pronoun but rather an invariant relative particle. Where the relativized noun is the object of a preposition, Welsh (but not Irish) has recourse to a resumptive pronoun within the relative clause --- exactly the same strategy used in Egyptian and (so Pokorny says) in Berber.⁵⁵ Here too, however, Pokorny (like Morris-Jones) sees what he wants to see in his Berber sources. He chooses to cite Stumme (Shilha dialect, 1899:96), where the only possibility given for prepositional relatives is

(a) the man that [I talked about him].

However, R. Basset (Kabyle dialect, 1886:20-21) attributes this construction to Arabic influence, giving by preference the more authentically Berber

(b) the city that [to I walked] "the city that I walked to", with preposed Prep and no prepositional object at all. And Hanoteau, whom Pokorny cites frequently elsewhere, makes no mention of the Arab-style construction at all (1896:36-38), but only of type (b), with the schematic example

(c) l'homme lequél [chez j'ai couché hier] .

Significantly, Hanoteau's grammar is of Tamashek (Touareg), the least Arabicized of all the Berber dialects. In present-day sources, neither Chaker 1983 (Kabyle dialect) nor Sadiqi 1986 (dialect intermediate

⁵⁵ Pokorny also states (1959:158) that the English "the man I talked to you about" self-evidently belongs to the same type and should be ascribed to Celtic influence --- a conclusion which ignores both the lack of resumptive pronoun in English, and the fact that such "dangling prepositions" also occur in non-relative contexts such as passives.

between Shilha and Tamazight) mentions the Arabic-style alternative.

14) Insular Celtic, far more than any other branch of Indo-European, makes heavy use of verbal nouns, which fulfill the functions of infinitives and participles in the other IE languages. The verbal noun can be construed either actively or passively (cf. "the shooting of the hunters") --- the sign of a "passivistischen Zug" in the language. A clear analogy can be found in the Egyptian "infinitive", actually a verbal nominal which governs its object in the genitive, and which likewise can be either active or passive. The putatively "insightful" parallelism with the prevalent ambiguity of valence in the Berber verb has been mentioned already, sec. 2.3.1.3.

15) Insular Celtic is of course hardly lacking in nominal compounding. However, compared to other Indo-European languages and even to Continental Celtic, it demonstrates a strong tendency to replace compounds by analytical genitive constructions, especially in personal and place names: Gaulish Rīgo-dūnum, Irish Dún Ríg. (For detailed exposition see 1927:367ff., 1955, and cf. Rhys 1890.) Such names involve such recurrent head-elements as fer (man), mac (son), áue (grandchild), cú (dog, wolf), mael, gilla (servant), muq (slave) (1927:370-72). The pattern is un-Indo-European, but echoes Semitic onomastics quite closely; cf. Hebrew names in ben (son), Ceved (slave/servant). Moreover, Irish also deviates from IE practice in names of the form "X son of Y": Gaulish onomastics has either "X + Y(gen.)" or "X + Y-yo-" (derived Adj.), but never "X son (of) Y" with "son" explicitly present as a separate word, a very common pattern in Irish as early as the ogam inscriptions (1955:57-58) and a normal type in Semitic, Egyptian, and Berber.

Pokorny even ventures an etymological link between Berber u (also au) "son" and Irish áue "grandchild, descendant" < IE *awyo- "grandparent"; only in Irish does this IE etymon mean grandchild, a circumstance suggesting a blend between the phonetically similar IE and Berber words (1955:59-60). --- This tendency to create analytic substitutes for old compounds is allegedly another "anreihend" trait.

16) Irish is said to have an aversion to adjectives (eine "deutliche Abneigung ... gegen den Gebrauch des Adjektivs"), a trait allegedly shared by Hamitic (cf. 1926:259ff., 1927:363-66). The statement depends on the implicit assumption (unspoken but inferrable from the discussion) that true adjectivehood inheres in the attributive, adnominal function of adjectives rather than their use as predicative elements; Pokorny's argumentation makes little sense without this dubious presupposition.⁵⁶ The main Irish datum presented in support of the stated "Abneigung" is the exclusively predicative nature of comparative and superlative forms in Irish (not "the biggest boy" but "the boy who is biggest"). Also adduced as evidence are genitival paraphrases of the sort "man of fame" (instead of "famous man") --- a common construction in Irish, but elsewhere as well, e.g. Latin. (Other minor points are mentioned briefly.) As arguments for Irish "adjectivophobia", these are marginal and not impressive; more convincing would have been data regarding the lexical frequency and the predicative-vs.-attributive behavior of non-compared adjectives in Irish, issues which are not addressed at all. Pokorny finds "parallels" all over Africa, including Bantu and "Negersprachen",

⁵⁶ Adjectivehood in this sense --- the Indo-European pattern --- is asserted, following Byrne, to reflect a more advanced form of linguistic expression (1926:259).

testifying (e.g.) to a paucity of adjectives and to the lack of a well-profiled system of adjective comparison in the Indo-European sense; these observations are valid (cf. Welmers 1973, chapter 9), but involve nothing directly comparable to the Irish phenomena. In Berber and Egyptian, Pokorny asserts (1926:259-60), true adjectives are rare and/or in competition with adjective-verbs ("to be-blue" instead of "blue"). But there is no indication of any particular rarity in the relevant sections of Gardiner's Egyptian grammar (1957:47, 108), nor (for example) in Chaker's grammar of Kabyle Berber (1983:101, 197-98); in both languages adjectives usually are deverbal and coexist with stative verbs, but this hardly amounts to an "Abneigung" or undermines their status as a vigorous and central part of the language. And again the phenomena are not comparable: in Irish, underived adjectives abound. The lack of strict comparability between the African and Celtic material adduced smacks of the "macro" typological thinking referred to earlier (sec. 2.3.1.3) --- appeal to a macro-dimension such as "degree of ideal Adjectivity", realized to different extents and in different ways in the various languages, any of which may be appealed to as comparanda.

17) When Insular Celtic deviates from its normal VSO order to yield SVO sentences, Pokorny asserts that the fronted element is in fact clefted ("so muss eine relativische Umschreibung stattfinden"), though immediately hedging his statement with the qualification that the copula often goes unexpressed ("oft nicht ausgedrückt"). "The king goes out" is thus more properly "(It is) the king (who) goes out". The same construction also exists in Egyptian and Berber. --- Pokorny's presentation here has the effect of explicitly conflating two distinct phenomena, clefting and topicalization. As such it diverges sharply from his treatment of 1927

(384ff.), where the Irish pendant nominative construction ("the king, he goes out") was presented straightforwardly as a fronted word-order variant with emphatic function --- a syntagm, again, with parallels in Hamito-Semitic (and in Indo-European).

18) The verb "do" plays a significant role in periphrastic constructions in Insular Celtic. In Middle Welsh it occurs coupled with the verbal noun ("he did VERB-ing"),

kyuodi a oruc "he arose"
arising PTCL he.did

a very common construction with an exact parallel in Egyptian (Gardiner 1957:395). Similarly in Irish, here with emphatic semantics (like English "do"+INF) --- though the Irish construction seems to me far less integral a part of the grammar than its Welsh counterpart. (Pokorny does not mention that analogous constructions are common in many other languages, e.g. Modern Persian.)

19) Insular Celtic lacks a verb "to have", instead using constructions like "is to-me"; likewise Egyptian and Berber (and Semitic). However, the construction is quite normal in older Indo-European (e.g. the Old Latin type mihi liber est "to-me is a book"), a point unmentioned by Pokorny.

20) The inherited Indo-European syntagm of nominal sentences is put to especially heavy use in Insular Celtic (cf. 1926:143-44). Elsewhere in Indo-European, Pokorny notes, nominal sentences occur especially in the 3rd person (cf. e.g. Meillet 1964:357) --- a restriction not found in Celtic (nor in Balto-Slavic). In Modern Irish such sentences are especially common in the "accompanying circumstance" usage with "and" ("and

my wife dead", see [2.vi] above); for older Irish he refers to Baudiš^v 1913a. Nominal sentences, of course, are a hallmark of Hamito-Semitic.

The above list mirrors the discussion in the 1959 article, but hardly exhausts the voluminous material Pokorny had at his disposal. Much more appears in his extensive treatment of 1926-29. We may mention by way of example the positioning and the non-repetition of the article in definite Noun-Noun embeddings in Irish: [Head the-Gen] --- just as in Arabic (1927:374). There is also a less insightful section (1927:384-93) devoted to the heavy use of the nominative case in Irish, not only as pendant nominative (see [17] above) but in a multiplicity of "non-IE" environments. Thurneysen (1946:155) characterizes the latter uses as occurring "where a noun stands in no precise syntactical relationship," an implicit and quite reasonable a priori appeal to the unmarked nature of the nominative; Pokorny (393) instead explains it as an attempt to render the caselessness of the substratum, but his African parallels appeal to Bantu as much as to Hamitic.

2.3.3 Reactions

Though numerous scholars have reacted to Pokorny's substratal thinking over the last half-century, only rarely has the reaction extended beyond a couple of brief lines in an article whose main focus is directed elsewhere.⁵⁷ The response has been overwhelmingly

⁵⁷ Exceptions are Baudiš 1926, Lansberg 1940, Bouda 1949. Note especially Greene 1966a, which provides perhaps the fullest and the most balanced summary account of pre-Celtic substratalism yet to appear; Meid

critical.⁵⁸ Only rarely, to be sure, do we come across open or veiled sarcasm; Binchy (1962:117-18) refers to "a somewhat exotic school of philologists", and C. Watkins (1962:1) to "a variety of hypotheses of the presence of sundry ill-defined non-Indo-European substrata in Celtic, particularly on the shadowy elfin-populated Emerald Isle."⁵⁹ Rare, too, are cases where Pokorny's facts per se are in dispute (notably Bouda 1949 on Basque and Caucasian) --- though, as we saw in sec. 2.3.2, factual errors or misrepresentations are not unusual. Rather, most criticisms express polite but firm reservations over various aspects of Pokorny's theory and methodology.

Several authors object to Pokorny's overzealous substratophilia per se. Thus Baudiš 1926, responding (in Czech) to Pokorny's attack (1926:130-132) on his earlier position (Baudiš 1923:120-126), endorses Pokorny's substratal inquiries in principle but reiterates his earlier methodological reserve: with regard to language mixture, "there are many unresolved problems, and I think that caution is necessary as long as

1972 also provides a balanced overview, but briefer (1193, 1196-98). Reactions after 1960 are of course directed at the work of Heinrich Wagner as well.

⁵⁸ Prior to Heinrich Wagner, the only unqualified endorsement of Pokorny's views which I have come across is Lansberg 1940, a brief and disappointing article. Bouda 1949, though critical of Pokorny's facts, endorses his overall thesis ("car somme toute il a bien raison", 336); Havers expresses approval of aspects of Pokorny's Substrattheorie, especially the notion of the delayed literary emergence of a suppressed stratum of folk speech (1931:138-39). Greene 1966a and Meid 1972 judiciously weigh both pros and cons.

⁵⁹ The two comments are unlikely to have been made independently of one another. On the very last page he ever wrote on pre-Celtic substrata (1964:80), Pokorny observes that Watkins was Binchy's student; "Es scheint," he says with some bitterness, "als ob in gewissen Kreisen in Dublin eine starke Abneigung gegen die Annahme nichtidg. Einflusses besteht."

these problems remain unresolved" (1926:217).⁶⁰ Bergin 1946:153, without further amplification, accuses Pokorny of "distort[ing] the evidence, arguing like an advocate, not like a judge." Sjoestedt-Jonval 1939:111, responding not to the Celtic-Hamitic problem but to Pokorny's substratal treatment of the Illyrian question, regrets his overenthusiastic appeal to the unknowns of prehistory --- "la méthode qui consiste à expliquer le connu par l'inconnu, et une assimilation phonétique par une forme de gobelet."

A number of critics take issue with Pokorny's contention that substratal influence manifests itself chiefly in syntax and phonology ("innere Sprachform"). This conclusion is unacceptable to Romanists (thus Spitzer 1928:443), and confuted by evidence from Rohlf's study of southern Italy (1930), where the Greek substratal influence shows itself instead in syntax and vocabulary (thus Dillon 1945:16; cf. also Havers 1931:138). Greene (1966a:125) makes a similar point with regard to Anglo-Irish; here the substratal influence is syntactic and phonetic but not phonological or lexical. Significantly, the one point of agreement uniting Pokorny and his various critics is that the substratum does have an effect on syntax.

Of the two basic methodological postulates of substratal explanation --- the importance ascribed to substratal influence per se, and the assumption that typological resemblances with other languages are historically significant --- Calvert Watkins alone of the critics appears

⁶⁰ My thanks to Gwendolyn Albert for her English translation of Baudis's Czech article. A capsule summary in German appears in Fraenkel 1928:361.

to take issue with the former. If an IE-internal explanation can be offered, he says, "then the necessity for recourse to such hypothetical substrata simply disappears" (1962:1).⁶¹ Most critics, by contrast, do accept the significance of substratal influence, but balk at Pokorny's uncritical use of the second postulate to establish the (Hamitic) identity of that substratum. Meid expresses this viewpoint perfectly:

... the notion of a strong substratum influence in Insular Celtic has a high degree of probability and should not be dismissed lightly. But ... Pokorny invites scepticism by trying hard to pin down the prehistoric peoples or cultures to which he thinks these particular influences are due.
(1972:1197)

Such skeptical reservations can sometimes be quite perfunctory --- thus Baudiš 1926:216 ("I do not identify the aboriginals of Ireland as to their ethnicity"), or Havers 1931:139 (the case for Hamitic "ist nicht erwiesen") --- but more often the critic presents an argument for his point. Some contend that the typological net must be cast wider. Spitzer (1928:443) presents this idea with great acuity:

Aber auch bei den Parallelen des Sprachbaus zwischen Hamitisch und Irisch hatte man das Gefühl, dass ebensogut andere Substrattheorien statt der Pokornyschen eintreten könnten: solche Zusammenstellungen könnten nur dann beweisend wirken, wenn ihnen ebenso liebevoll zusammengestellte Listen von Übereinstimmungen des Irischen mit anderen Sprachfamilien ...

⁶¹ Recall the similar views expressed by Martinet, see sec. 2.3.1.1.

gegenübertraten und dann das Hamitische den Rekord schlänge.

[But even with the structural parallels between Hamitic and Irish one gets the feeling that other substratal theories could have been appealed to just as well as Pokorny's: such [Celtic-Hamitic] compilations would be probative only if they were to be juxtaposed with arbitrarily compiled lists of agreements between Irish and other language families ... and Hamitic then came out the winner.]

Hubschmid, himself an avid Mediterranean substratalist (working primarily with lexical materials), rejects Pokorny's conclusions on similar grounds: an examination of other, different non-IE families would doubtless yield other points where Celtic deviated from IE but agreed with the given non-IE language (1960:40). Finally, Vendryes issues a methodological caution against the ascription of historical significance to typological resemblances, pointing out that independent convergence should not be discounted: "Les possibilités linguistiques sont en nombre limité: les mêmes motifs d'évolution aboutissant aux mêmes résultats se retrouvent en des langues diverses, qui n'ont entre elles aucun lien historique" (1937:345; the passage does not mention Pokorny by name, but the inference is clear).

Other critics focused instead on the legitimacy of certain of Pokorny's "typical Celticisms", Celtic peculiarities which he presents as inexplicable in purely IE terms. Upon closer inspection, these might turn out after all to recur elsewhere in IE, perhaps not in literary prose but in everyday vernacular speech, and notably in languages where no Hamitic substratum could possibly have played a role. Baudiš

(1926:217) applies this criticism to Pokorny's claim that the short choppy sentences typical of Irish narrative prose are something distinctively Celtic and non-IE. Bergin (1938a:214) makes the same point with regard to the special Irish fondness for pendant nominatives. In this connection Thurneysen (1930:428) appeals for an anti-Pokornyan "Gegenadvokat", who should "genau untersuchen, ob die einzelnen Erscheinungen nicht auch bei andern Indogermanen vorkommen, die keines hamitischen Substrats verdächtig sind" (1930:428). He suggests further that Irish prose largely escaped the pervasive Greek and Latin stylistic molds that served as models elsewhere in medieval Europe, a fact which might explain --- in literary, non-linguistic terms --- certain aspects of its "non-IE" cast. (See Pokorny's reaction, 1949:235, as discussed in sec. 2.3.2 [2, vii] above.) In the same general vein, though with a very different particular focus, Dillon and C. Watkins cast doubt on the special "Celticity" of the structure of the Irish verb, appealing to parallels with Hittite (Dillon 1947:24) and presenting the Irish verb as a natural evolution from Indo-European (C. Watkins 1962, 1963). (For Pokorny's response see his articles of 1949, 1964, and sec. 2.3.2 [12] above.)

Other methodological criticisms appear as well. Vendryes insists that any valid substratal work must of necessity be preceded by an exhaustive attempt to account for the phenomena in orthodox comparative terms (1937:345). Mac White, discussing not only Pokorny but other substratal proposals regarding Celtic, points to a certain geographic fuzziness attendant on the aggregate of such hypotheses. Two questions, he says, require clarification: what exactly is to be attributed to Hamitic and what to Basque; and whether the substrate operated only in the

British Isles or also on the continent (1955:10-12). Finally, Greene criticizes Pokorny's indiscriminate affection for any and all substratal links, observing that he has "weakened his case by seeking parallels in too great a number of languages" (1966a:127).

2.4 Heinrich Wagner

A warm personal and professional bond linked Julius Pokorny and Heinrich Wagner, the third leading advocate of the Celtic/Hamito-Semitic connection. Pokorny and Morris-Jones, though on cordial terms,⁶² had belonged to two different cultural and intellectual worlds. By contrast, Wagner was Pokorny's disciple, and his orientation and oeuvre can justly be taken as an organic outgrowth of Pokorny's. Wagner's scholarly career, 1950-1988, overlapped Pokorny's by two decades; indeed, Pokorny's later work (e.g. Pokorny 1960a:238, 1960b) acknowledges its debt to Wagner's Das Verbum, a book dedicated in turn to "meinen verehrten Lehrern und Freunden Julius Pokorny und Ernst Lewy".

Though the Celtic/Hamito-Semitic connection dominates much of Wagner's oeuvre, only a single article is expressly devoted to the problem (1981), a more or less non-technical presentation for the general audience. By far the most comprehensive treatment is embedded in his 1959 book, Das Verbum in den Sprachen der britischen Inseln, notably sections III.A and D (pp. 152-82, 205-40); but the Celtic/Hamito-Semitic problem, only one of many themes explored in this multidimensional book, is not easily disentangled from the remainder. Also of primary

⁶² Witness Pokorny's letters to M-J, cited in S. Zimmer 1986:283-84.

importance is the 1976 study of language in pre-Celtic Britain and Iberia; and, to a lesser degree, the articles of 1972a (secs. 2-7) and 1987 (on Celtic migrations to the British Isles). The articles of 1967 and 1977 (also parts of 1972a) address the archaicity of Old Irish within Indo-European, arguing against old linkages between Celtic and Hittite (and sometimes in favor of linkages between Celtic and the Mediterranean). And the areal studies of 1951 (Germanic; very brief), 1959 (Appendix on Indo-European), 1964 (North European "Lautgeographie"), 1969 (Continental Celtic), 1978 (ergativity), and 1985 (Hittite) provide insight into Wagner's areal view of prehistory and his linguistic Weltanschauung and methodology.

Wagner's approach to language was militantly "anti-algebraist", anti-theoretical, organic, and humanistic: "Sprachen sind nicht zweckmässige mathematische Systeme, sondern äusserst verwickelte, historisch gewachsene geistige Organismen" ["Languages are not teleological, mathematical systems, but historical products, 'geistig' organisms of the greatest complexity"] (1964:284). This iconoclasm encompassed both synchrony and diachrony. From a synchronic perspective, Wagner

tended to think that taxonomic structuralism, and its further elaboration by Chomsky and his followers into generative grammar, was an aberration. He was against deductive theory and refused to accept any view of language which forced him to follow narrow rules and regulations. (Mac Mathúna 1989:217 [necrology])

To the "spekulativ-strukturalistische" argument (1964:283) that gemination in Old Irish is nondistinctive, for example, Wagner counters from a

contrastive and areal perspective (284): "Es sind gerade die phonologisch irrelevanten Züge eines Lautsystems ... welche Sprachen charakterisieren und voneinander unterscheiden" ["It is precisely the phonologically irrelevant features of a sound system [e.g. nondistinctive gemination] ... that characterize languages and differentiate them from each other"].⁶³ He was equally hostile to diachronic orthodoxy, as embodied in the "mechanistic" approach of the neogrammarians:

Die phantastische, unrealistische Auffassung der philologisch-junggrammatischen und der modernen strukturalistischen Indogermanistik, Sprachgeschichte beruhe im wesentlichen auf innern Veränderungen eines sprachlichen Mechanismus, der in Grunde nichts mit den Völkern und deren Geschichte zu tun habe, muss ich kategorisch zurückweisen. (1964:264-65)

[The fantastical, unrealistic view held by Indo-Europeanists of both the philological-neogrammarian and the modern structuralist school, whereby language history in essence is based on internal changes to a linguistic mechanism that has fundamentally nothing to do with peoples and their history, I must categorically reject.]⁶⁴

⁶³ Yet in practice Wagner could be a very good structuralist. Thus he emphasizes that, to understand the function of particular forms in a language, the linguist must look at "die bedeutungsmässigen und funktionellen Oppositionen" inherent in the system as a whole (1959:183). The first half of Das Verbum does precisely this for the verb systems of the British languages.

⁶⁴ Wagner was not shy about expressing such antiestablishment views. They form a subtext running through all his writings, a subterranean stream typically bubbling to the surface through the footnotes. The 1964 article is especially outspoken in this respect.

Grammatical categories must not be treated purely from a "Darwinian" perspective, as somehow simply "entstanden" through purely internal means such as sound laws, analogy, speech errors, "Systemszwang", etc. (1972a:68, cf. 1972b:316) --- indeed, "die 'Erklärung' einer grammatischen Kategorie aus einem 'Lautgesetz' [ist] a priori als falsch zu bezeichnen" ["the 'explanation' of a grammatical category [as arising] from a 'sound law' is a priori wrong"] (1959:85, cf. 1967:305). Like Pokorny, Wagner does accept the traditional notions of genetic relatedness and protolanguage; (Proto-)Indo-European was surely "eine einheitliche Sprache im Sinne der Stammbaumtheorie" (1959:241), and in general the word-forms of language ("Materie") are legitimately to be treated in terms of orthodox historical linguistics (242). But this will not yield much insight into the evolution of grammatical categories, of overall language structure. Here "typological relatedness" and geography instead come to the fore (see sec. 2.4.2.1 below).

Wagner's writing poses a challenge to even the most sympathetic reader. The erudition, exotic data, and conceptual subtlety are embedded in a rhetorical and terminological matrix all their own, and one which can be bafflingly opaque --- "mehr verschleiernd als erhellend", as Rössler put it (1960:142).⁶⁵ The presentation must be not so much followed as unraveled. Like Pokorny, Wagner often argues in an "anreihend" style, through the associative chaining of ideas or examples rather than the logical development of a theme.⁶⁶ The typological

⁶⁵ A single example (which has not been noted elsewhere): such phrases as "das grammatisch bestimmende Element" (1959:81, 145, 156, 253) mean not "the determining element" of a construction (Dependent) but "the governing element" (Head) --- backwards from normal usage.

⁶⁶ Nothing short of lengthy verbatim quotation could truly convey the

comparison is heavily theory-bound, and the theory is far from transparent and often far from plausible. There is a readiness (verging on eagerness) to take homonymy as identity --- notably with the various d morphs in Berber (1959:173, 175-76, 178-79, 212) and in Irish (170-73), which are freely equated not only intra- but inter-lingually. Finally, opposing points of view, when brought up at all, are typically mentioned en passant and only for brief dismissal: "Die Mühe des Gegenbeweises hat sich W[agner] nicht gemacht" (Schmid 1960:315).

2.4.1 Typology: terminology and constructs

The typological analysis in Das Verbum rests principally on two typological macro-dimensions: nominal vs. verbal, and anreihend vs. unterordnend. For Wagner, verb syntax/semantics has both a "nominal" component --- the concrete lexical meaning --- and a "verbal" component --- the verbal action per se ("Aktion, Handlung"), encompassing the verb's potential inflectional categories: tense, aspect, mode, person (1959:xii-xiii, 5). Languages differ greatly in the relative importance they assign to these components, so that there are both "nominal" and "verbal" languages (technical terms in Wagner's lexicon). "Nominal" languages allow much lexical nuancing in the inflecting verb stem, through rich derivational morphology. In highly "verbal" languages, by contrast --- such as Insular Celtic, Basque, and Hamito-Semitic --- the finite inflecting verb tends to be semantically impoverished. Some languages of this type will operate with verb roots rather than verb

fluid (and frustrating) quality of this kind of exposition; see e.g. secs. 2.4.3 [6.v and 7] below.

stems, with little in the way of meaning-changing, "nominal" derivation whether prefixal or suffixal; Wagner refers to such languages as "wurzelflektierend" (root-inflecting). Or the language may make heavy use of inflecting finite auxiliaries (conveying the pure abstract fact of verbal action) in combination with a subordinated verbal noun (conveying the lexical meaning), yielding periphrastic tenses of the general form "he is at VERBing", "he does VERBing", etc. Here the meaning-bearing (nominal) part of the verb is totally severed from the inflectional (verbal) part, a phenomenon Wagner terms "bedeutungsisolierend" (meaning-isolating) and occasionally, following Lewy (1942:26), "flexionsisolierend".⁶⁷ Basque is the prototypically meaning-isolating language; the modern Brythonic Celtic languages come close. Older Celtic languages and Hamito-Semitic tend to be root-inflecting; Indo-European, by contrast, is "stem-inflecting".

The long-standing opposition of "anreihend" vs. "unterordnend" languages undergoes a sea change in Wagner's hands. The issue is no longer morphological word-type, or the tight/loose integration of morphemes in the word (recall sec. 2.3.1.3). Rather, word order --- originally regarded as an epiphenomenon, a contingent and secondary feature of these macro-types --- has now become criterial and definitional. "Anreihend" is defined (1959:156) exactly as we would define the "ideal VSO" macro-type; SVO is a variant of the "anreihend" type (159); "unterordnend" is the "ideal OV" type.⁶⁸ Uralic and Altaic are the

⁶⁷ The two terms seem tautologically synonymous; cf. the pleonastic reference to "Flexions- und Bedeutungsisolierung" (1959:102). For areal discussion of Bedeutungsisolierung, see 1959:98-103. The terminology belongs to the Humboldtian typological tradition (Finck, Lewy).

⁶⁸ Note that these characterizations predate Greenberg's study of word

prototypical "unterordnend" languages; prototypical "anreihend" languages include not only VSO Insular Celtic and Hamito-Semitic, but also SVO Bantu (209).⁶⁹

The opposition of VO vs. OV word order and the opposition of verbal vs. nominal languages would seem to be independent notions. But in fact, in the languages under consideration, the two show a strong (though not universal) correlation. Ural-Altaic is both OV (unterordnend) and nominal; Hamito-Semitic and Insular Celtic, both VO (anreihend) and verbal. Thus Wagner frequently speaks of a Eurafrikan "anreihend-verbal" type, over against the Ural-Altaic "unterordnend-nominal" type which is its polar opposite (1959:182, cf. 1969:208). At one point, in fact, this is presented as a natural coupling: "Die Sprachen des anreihenden Sprachtypus sind verbale Sprachen: Die Aussenwelt wird vom Standpunkt der Handlung ... aus beschrieben" ["Languages of the 'anreihend' type are verbal languages: The external world is described from the perspective of the action"] (1959:252). Basque is a problem in this schema, being simultaneously very "anreihend/verbal" yet primarily "unterordnend" (e.g. 1959:101).⁷⁰

order (1963, published 1966); the notion of ideal VO and OV macro-types has been part of the European typological tradition for at least a century. The multi-factor nature of these definitions is quite important to Wagner. In his discussion of Old Irish word order and of "archaic" OV constructions (e.g. 1977:208), he criticizes others for focusing only on the position of the verb rather than on the whole constellation of OV/VO word order factors involved in poetic alliteration. But there are slips. Wagner repeatedly presents Akkadian (with clausal SOV order) as basically an "unterordnend", OV-type language --- and as such comparable to Ethiopic Semitic (1976:402) and to Hittite (1985:17) --- whereas in fact Akkadian is transparently an "anreihend", VO-type language in every respect except clause-final position of the verb (Gensler 1989).

⁶⁹ In fact, SVO Bantu is an extremely good "anreihend" language; Wagner calls it "exzessiv anreihend" (1959:253).

⁷⁰ We will see that Wagner's treatment of Basque is self-contradictory

One of the knottiest and yet most important aspects of Wagner's treatment of the "anreihend-verbal" type (both Berber and Insular Celtic) is the reanalysis of pre-verbal particles as themselves being verb-like.⁷¹ This would be in keeping with both the "anreihend" and the "verbal" nature of the language: the particle, being in absolute sentence-initial position, has the position which by rights belongs to the verb in an "anreihend" (VSO) language; further, to recast the particle as an abstract helping-verb of sorts, with the main verb subordinated to it, is very much in keeping with the essence of a "verbal" language (meaning-isolation). Such an analysis comes down to an assertion that the pre-verbal particle (NEG, for example) acquires predicational force. Wagner does not express the idea this way, but says that the particle takes on a "copular" character --- as if the only way the notion "NEG" could acquire verbhood were to recast it as "(BE) NEG". The tangibility of this copula varies: sometimes the particle itself is recast as a copula, sometimes it includes a copular morpheme (or the historical residue of one),⁷² sometimes it contains a zero copula or a "virtual" copula (thus the Irish NEG prefix ní- is "mindestens virtuell kopulahaltig" [1977:211]). The copula, moreover, is impersonal, i.e., devoid of person/number marking (1959:13, 169); such an impersonal copula is patent in Modern Irish (is), and Wagner sees it also (inter (sec. 2.4.2.3 below).

⁷¹ Thus preverbs in modern Irish "sind eben selbst quasi-Verben. ... Jede von einem Praefix eingeleitete Verbalform ist im Grunde genommen grammatisch untergeordnet" (1959:221). Much of what follows here is my own attempt to make general sense of Wagner's impressionistic presentation; for detailed critique see sec. 2.4.3 [6] below.

⁷² Wagner chronically fails to distinguish between the dead diachronic residue of a morpheme and the living synchronic presence of the morpheme itself --- a point we will return to (sec. 2.4.2.2).

alia) in Old Irish d and Berber d (see sec. 2.4.3 [6] below).

Though Wagner is seldom explicit about the syntactic structure underlying this alleged "verbal-preverb" approach, there are three logically distinct construction types which such an analysis might cover:

1) The abstract Head verb is inflected as a helping verb, with the concrete Dependent verb subordinated to it as an non-inflecting verbal noun. Thus there are languages (Finnish, Egyptian (tm), even English to a degree) where the ordinary, unmarked negation of "he bears" involves a conjugated "negative verb" ("he NOTs to bear"); similar is the Hebrew construction whereby e.g. "he sings a lot" is rendered literally as "he muches to sing". Of the three types, this is conceptually the clearest: the concrete verb (Dependent) does not inflect like a verb, whereas the abstract helping verb (Head) does, lending the analysis a palpable reality. But neither this nor the next type figures in Celtic or Berber.

2) Both concrete Dependent and abstract Head are normal inflecting verbs.

3) Only the concrete Dependent inflects; the abstract Head is invariable (impersonal). This type is central to Wagner's analysis of the Irish and Berber verb (invariable preverb, inflecting main verb). At the same time, of the three types this is surely the one where reanalysis of the preverb as a "verbal" element is hardest to motivate; for it is not the preverb but the ordinary main verb which inflects like a verb.

Conceivably it was an awareness of this difficulty that prompted Wagner's recasting of such "verbal preverbs" as impersonal copulas, for a copular analysis immediately suggests a straightforward and familiar

interpretation: as a mise en relief or clefting construction (e.g. 1959:173-75). Thus "NEG he-bears" would be

(It) (is)-NEG (that) he bears

i.e. "It is not (the case) that he bears".⁷³

This interpretation is corroborated by Wagner's gloss on the example

nicon-chloor "let me not hear" [with NEG preverb nicon-],

analyzed as "nicht-ist-dass höre-ich" (1967:304-5; cf. Greene's criticism, 1969). He explains further (1977:211) that, in Old Irish (and Egyptian), sentences with negative and interrogative preverbs are to be taken on a par with positive "mise-en-relief" (cleft) sentences; schematically:

Neg: Nicht ist es, dass er kommt

Ques: Ist es, dass er kommt?⁷⁴

Cleft: Er ist es, der kommt = c'est lui qui vient.

But there are major analytical problems here. Both Irish and Berber have special relative forms of the verb; yet in these Preverb-Verb constructions, allegedly clefts, neither language uses the relative form of the "embedded" (main) verb. Further, the notion "emphatic" (mise-en-

⁷³ Meid (1963:51) objects that such an analysis runs counter to one's straightforward Sprachgefühl: Old Irish ní beir means simply "he does not bear", not "it is not that he bears". But Wagner (1967:305) rightly takes Meid to task for such a "Sprachgefühl" argument. One could point out that in those languages (above) where the negative is expressed by a conjugated "negative verb" ("he NOTs to bear"), the syntax doubtless also violates Western European "Sprachgefühl"; yet this remains the ordinary, unmarked negation of "he bears".

⁷⁴ French "est-ce que" interrogatives are of the same type (211). The example is especially instructive: synchronically, French [eskə] is an invariant preverbal particle; it contains the residue of a copula; hence for Wagner it must itself have copular force, and the sentence counts as a "mise-en-relief". See sec. 2.4.2.2 below.

relief) ought to make sense only as part of an opposition with a non-emphatic counterpart, yet these constructions (being the ordinary interrogative and negative) have no such counterpart.⁷⁵ The multiple confusion attendant on this treatment of clefting is typical of Wagner's intuitive approach, involving problems with syntax (copula? relative clause?) and semantics (emphasis?) and even diachrony (frozen copular residue versus "live" copula).

None of this speaks well for the analysis of preverbs as quasi-verbs. It is unclear whether a clefting interpretation is even possible. It is still less clear what other interpretation might replace it.

It should be stressed, finally, that "verbal" and "anreihend" are clearly macro-typological constructs; I illustrate here only for the latter. At various points the following features, logically independent of word order, are presented as "typisch anreihend" (or the like):

- a) The fundamentally punctive character of the finite verb (1959:146).
- b) Clefting ("mise en relief") constructions (1959:173).
- c) A close semantic (1959:179) and grammatical (1967:306) link between copula and relative marker, between copula and 'and', and between copula and deictic marker (1959:253, referring to pp. 175-76).
- d) A pervasive grammatical principle of "durchbrochener Parallelismus" ("anti-parallelism") whereby two conceptually parallel syntactic units are treated in grammatically unparallel ways (1959:207-40); see sec.

⁷⁵ As already remarked (sec. 2.4), Wagner had little use for such structuralist arguments. I suspect that, for Wagner, the emphatic/clefting nature of these constructions had little to do with notions of opposition; rather, it was simply in the nature of "anreihend-verbal" languages to express such things emphatically.

2.4.3 [10ff.].

e) Gender and nominal classification (252-56).

f) The "Satzwort" character of words (205); see sec. 2.4.3 [21] below.

2.4.2 Sprachgeographie

2.4.2.1 Wagner as linguistic geographer

Perhaps the dominant theme of Wagner's linguistic career was "Sprachgeographie" or "linguistic geography", the study of "Sprachlandschaften" ("language landscapes/areas") and "Sprachschichten" ("language strata") --- areal linguistics of a very particular stripe. Areal investigation for Wagner ran the gamut from the microscopic to the truly macroscopic, in the process ranging over a variety of linguistic sub-disciplines not normally thought of as belonging together. On the micro level is his massive study of dialect geography in the classical mold, the four-volume Linguistic Atlas and Survey of Irish Dialects (1958-69). On a slightly broader scale is his comparative typology of the Insular Celtic verbal system(s) as a whole (1959 [Part I]), a study encompassing all the Insular Celtic languages and focusing especially on the analytic, periphrastic re-formation of the synthetic tense-forms which early Celtic had inherited from Indo-European: the considerable variation displayed here by the Insular Celtic dialects shows a striking geographical progression from west to east in increasing degree of analyticity. Wider still is the scope of two studies devoted to the areal status within the Indo-European family of particular Indo-European sub-groups, namely Continental Celtic (1969) and Hittite (1985): each group is presented as dovetailing structurally with its present or past

geographical neighbors. Similar in scope is his 1964 areal analysis of the phonetics/phonology of the languages of northernmost Europe: Lappish, Icelandic, Scots Gaelic. And finally, on near-global scale, is the issue of pre-Indo-European "Schichten" in the British Isles, with comparanda ranging from westernmost Europe to the Caucasus and East Africa (1959 [Part III], 1976, 1981).

For Wagner these disparate types of investigation stand on a par, as co-equal manifestations of the same guiding principle: that similarities and dissimilarities between languages must make sense in geographical terms.⁷⁶ Such geographical distributions may be lexical, as in traditional dialect geography, or typological (1969:203); and though Wagner worked on linguistic geography in both of these senses, he explicitly situates himself (1976:395) in the latter camp, that of structural/typological Sprachgeographie. This approach was not new with Wagner; the passing reference to "die Energeia der 'Sprachlandschaft'" (1964:248) evokes the long Humboldtian tradition he was heir to. His areal orientation draws on Trubetzkoy's famous article of 1939, in which the typological intermediacy of Indo-European vis-a-vis Finno-Ugric and Caucasian was explained with reference to a parallel geographical intermediacy (similarly Wagner 1959, Appendix). More directly, it perpetuates the areal linguistics of Wagner's teacher Ernst Lewy, who unfailingly attached great explanatory importance to linguistic areality both synchronically and diachronically: "Wir sehen immer und überall, dass ... gewisse Eigentümlichkeiten der Sprachen 'am Boden haften'" ["Always

⁷⁶ "Jede Sprache ist mit ihrer Nachbarsprache typologisch verwandt" (1959:241); "Der Typus einer Sprache ist durch deren geographische Lage bestimmt" (1959:102); and so repeatedly.

and everywhere, we can see ... how certain characteristics of languages 'stick to the ground'] (Lewy 1942:26). The phrase is a kind of slogan for linguistic geography, and Wagner echoes it passim (e.g. 1959:178, 1972a:48).

At first glance this approach is simply a version of Sprachbund linguistics. But the claim is stronger: geographical closeness always implies typological affinity --- or "typological relatedness", to use Wagner's phrase. Stronger still, the one-way implication sometimes becomes a two-way statement: a typological connection implies a geographical one. And for typologically similar languages which are in fact not in contact, and which may even be rather far apart, the geographical connection is a reconstructed one that applies at an earlier stage in their history. In practice this involves either positing original geographical contiguity of the two speech communities involved, followed by migration of one of them (thus for Wagner's 1969 analysis of the original homeland of the Celts), or else assuming an earlier linguistic area, originally embracing not only the two speech communities but also the intervening territory, which was later partially overlaid by languages of another linguistic type (thus with the Celtic/Hamito-Semitic parallels).⁷⁷ It is unclear how far Wagner would have been willing to push this reverse principle --- presumably the

⁷⁷ Another example of the second type: Caucasian languages, Sumerian, and Burushaski all have OV sentence structure and relatively polysynthetic verbs; therefore "müssen wir annehmen, dass es zwischen Kaukasus, Kleinasien und Indien in vorhistorischer Zeit noch andere Sprachen dieses Typus gegeben hat" (1972a:38). Elsewhere Wagner justifies his "kaukasisch-kleinasiatisch-vorderasiatischen Sprachbund" by invoking structural links reaching all the way to Burushaski and Tibeto-Burman (1977:205-6).

shared OV type of Turkish and Quechua would not be explained geographically⁷⁸ --- but he certainly applies it in concrete cases.

The linguistic Weltanschauung of the ideological substratophile is powerfully in evidence here, even if in recast form. Sprachgeographie becomes the counterpole to the "algebraic" current in linguistics which Wagner so much disliked, and as such it is invested with all the potency of a linguistic touchstone:

Unsere Untersuchung geht dementsprechend vom Raume aus und von der Erkenntnis, dass geographische Räume für die typologische Ausgestaltung konkreter Einzelsprachen von entscheidender Bedeutung sind. Während die philologisch-indogermanistische Methode der Rekonstruktion luftleere Räume schafft, stellt unsere Methode den Sprachraum selbst in den Mittelpunkt...

[Our investigation, accordingly, begins with space/area, and with the recognition that geographical areas are of decisive importance in the typological development of concrete individual languages. Whereas the philological-Indo-Europeanist method of reconstruction creates vacuum spaces, our method places the linguistic space/area itself at the center...]
(1964:231-32).

For Wagner the tangibility of geography imparts an organic vitality and validity to any linguistic analysis done in the "sprachgeographisch" spirit, in contrast to the stifling vacuity of traditional historical

⁷⁸ Wagner almost never discusses New World languages, even in his typological study of ergativity (1978); elsewhere (1976:406-7) the New World is cursorily disposed of as "Amero-Indian".

linguistics.

But the weaknesses of the method should be plain. Wagner's thinking embodies an all-too-apparent mental blank regarding the legitimacy of "coincidence" as a real concern in linguistic typology.⁷⁹ And geography, though an extremely important determinant, is demonstrably neither necessary nor sufficient as an explanation of typological distributions. Of course, where a geographical connection between structurally similar languages is patent, i.e., can be asserted independently of the linguistic data --- adjacency or near-adjacency of speech communities, or a documented commercial or political or migratory link between non-adjacent lands --- then geography, as an independent variable, has great prima facie explanatory value. By contrast, if a proposed geographical connection is not a self-evident fact but rather must be argued for on the basis of existing distribution of linguistic structures (as with a posited ancient linguistic area, now attested only in a scattering of languages), then a geographical explanation of that distribution can involve an element of circularity. In such cases, Wagner's decision to undertake an areal analysis at all verges on an act of faith. Methodologically the two cases are poles apart, and involve very different degrees of reliability. Yet Wagner's approach conflates them as a matter of principle.

The theoretical grounding Wagner provides for his sprachgeographical approach (1959:241-42) is profoundly Humboldtian in its orientation,

⁷⁹ Thus his sarcastic dismissal: "das Wort 'Zufall', ein bei junggrammatischen und strukturalistischen Sprachforschern recht beliebtes Wort" ["the word 'coincidence', a word neogrammarian and structuralist linguists are awfully fond of"] (1964:295).

and by that very token profoundly confusing. Language consists of concrete "Materie" (the stuff of language, its "Wortformen") superimposed on deeper "innere Sprachform". Traditional historical linguistics, dealing with language as "Materie", has arrived at a genealogical classification dependent on the social history of speech communities (conquests, migrations, etc.). By contrast "innere Form", as expressed through the given language's system of linguistic categories, is determined anthropologically, by man's "Geistesstruktur" (or "Menschentypus"); in Das Verbum this is presented as a matter of natural history.⁸⁰ Since different peoples have different Geistesstrukturen, they will naturally speak languages having different "innere Form". But Geistesstruktur is not linked to any particular "konkretes Volk"; rather, it is "geographisch bestimmt" (242). As neighboring peoples mix and intermarry, their Geistesstrukturen and hence the "innere Form" of their languages will take on structural similarities, inevitably leading to the geographical continuum of structural types which lies at the heart of Wagner's "sprachgeographisch" approach. Similarly, when a language family disperses, each daughter language will naturally gravitate toward the structural type characteristic of its new geographical home.

A major source of confusion in this presentation is the attempt to cast Geistesstruktur as part of natural history, as something inherently

⁸⁰ "Typologische Sprachverwandtschaft beruht m.E. auf anthropologischer Verwandtschaft ...; sie folgt nicht der Sozialgeschichte des Menschen, sondern seiner Naturgeschichte." Thus in 1959. But compare the following socially oriented statement from 1972: "Typologische Sprachverwandtschaft hängt m.E. mit der menschlichen Sozialgeschichte und der Verteilung von Völkern und Rassen zusammen" (1972a:68). The two viewpoints are inverses.

bound to geography. This may emerge from a desire to make Sprachgeographie "scientific", or may reflect the tradition of correlating language type with race (thus Finck, or Lewy [e.g. 1961[1951]:9-10]). But how would such a natural-historical account work? Wagner does not say. Presumably he would not go so far as to invoke an "overmind" hovering over a region like a kind of local spiritual weather system. Perhaps he felt Geistesstruktur to be part of the human genetic endowment, transmitted --- within and across populations --- through the biological process of reproduction. A gene-type is (normally) geographically localized and geographically stable; hence Geistesstruktur would be at once geographically determined and a part of natural history. Of course the objections to this view are well known (children easily learn any language they are exposed to); Pokorny himself had repudiated it as far back as the thirties (1936:70; see 2.3.1.2 above). To read it into Wagner's thinking in 1959 is perhaps improbable, but little else suggests itself.

Correlations between linguistic type and national mentality (Geistesstruktur, Menschentypus, linguistic Weltanschauung) are taken for granted throughout Das Verbum, less explicitly so in Wagner's later works (see e.g. 1959:44, 50; 1978:57). They occur especially in connection with the "anreihend" and "unterordnend" macro-types, thereby anchoring the linguistic types in extralinguistic reality. Thus the "unterordnend" (OV) type of Finno-Ugric manifests a (subconscious) linguistic Weltanschauung whereby "die Aussenwelt wird nicht vom Standpunkt der Aktion aus betrachtet, sondern sachlich-nominal" (1959:145); by contrast, the Berber verb belongs to the "anreihend-verbal" (VO) type, which conceptualizes "die Aussenwelt eher vom

Standpunkt der Handlung als vom Standpunkt der Sache" (205). Such correlations, however, are intended to be value-free; no language or nationality is "better" than any other, no language is "primitive". Wagner condemns "das Märchen vom parataktischen Charakter der 'Primitivsprachen'" (215); similarly, it is wrong to take the lack of a Biblical Hebrew word for "yes" as manifesting an early linguistic stage "wo der abstrakte Begriff der Bejahung noch fehlt" (223-24).

A particularly subtle manifestation of Sprachgeographie lies in the typological/geographical links that Wagner posits at the etymological level --- the notion that "zwischen Sprachen, die nicht genealogisch, sondern nur typologisch-geographisch miteinander verwandt sind, auch etymologische Beziehungen bestehen" (1959:178). The issue concerns areally linked languages which have a grammatical morpheme showing close agreement in both form and function. Wagner's parade example is the grammatical -d- element of Irish, Berber, Basque, and Fula, all of which are considered related in a certain sense: "Kelt. d [ist] mit berber. d verwandt, nämlich lautlich und syntaktisch" ["Celtic d is related to Berber d, namely phonically and syntactically"] (178).⁸¹ Other examples are Hittite/Akkadian -man (conditional/irrealis) (178), Hittite lē/Akkadian la (negative) (178), Modern Irish is/Anglo-Irish it's in such constructions as "it's flat it was" (178), Modern Irish bí/Anglo-Irish bees (habitual) (113), Berber m-/Old Irish imm- (reciprocal) (199), Basque ba/Welsh pe (irrealis) (230), and English she/Irish sí (1972a:69). Not only does Wagner consider these agreements non-

⁸¹ Similarly: "In der Tat haben jedoch gewisse grammatische Elemente eine geographische, nicht an genealogische Verwandtschaft gebundene Verbreitung" (1967:305; cf. also 1972a:68-69).

coincidental;⁸² he also rejects the straightforward explanation, borrowing. Such asemanic, grammatical elements, he explains, are not normally loaned like concrete vocabulary, but instead "schliessen sich nun aber offenbar eher den syntaktischen Beziehungen zwischen zwei Sprachen an und ... haften wie syntaktische Erscheinungen am Boden" ["plainly are associated, rather, with the syntactic links between two languages and ..., like syntactic phenomena, stick to the ground"] (1959:178). In context, with the words "innere Sprachform" occurring two lines earlier, the intended association is clear: grammatical morphemes count as part of "innere Sprachform" and as such are to be approached from the same sprachgeographical perspective.

There is a good insight lurking here, but expressed poorly.

"Innere Sprachform" is not the issue: Wagner's own words make it clear that this term has to do with systems, with a language's "lautlichen und syntaktischen Bau" (178), with "Lautsystem und Syntax, weniger Morphologie und Wortschatz" (242). A concrete sound-meaning unit, grammatical element or not, has nothing to do with this. Further, a total rejection of grammatical borrowing is too strong (on this see e.g. Weinreich 1963:29ff.). However, in a bilingual speech community it is thoroughly plausible that a morpheme in L1 and a morpheme in L2, originally unconnected and showing only vague resemblances in sound and meaning, could converge and grow increasingly similar over the course of time. To be sure, it would seem that both lexical and grammatical morphemes should

⁸² Coincidence, of course, is quite possible: note Imbabura Quechua -man (conditional) vis-a-vis Hittite/Akkadian -man (above) (Cole 1982:14), or the the negative imperative morpheme al found in both Hebrew and Ineseño Chumash (California) (Applegate 1972:329).

be equally susceptible to such a process. But a convergence that specifically involved a grammatical morpheme might well proceed in tandem with, and be fostered by, a convergence in the overall grammatical systems of the two languages (of the sort described for the village of Kupwar, India, by Gumperz and Wilson 1971). This I take to be the insight Wagner is groping at. --- The key to this scenario is the notion "speech community", a concept that becomes less and less tenable the greater the geographical distances involved. Thus Wagner's Akkadian/Hittite and Anglo-Irish examples, involving Sprachgeographie at the micro-scale, are at least plausible; the case of -d-, involving a scattering of locations over vast geographical distances, is far less so, and similarly for the Berber/Irish reciprocal m.

A final point, in a rather different vein. One of Wagner's recurrent concerns is to reconstruct the proto-geography of a group of languages on the basis of their present location and of the distribution of features found to varying degrees in the different languages. Any such attempt must come up against the classical problem of "marginal" vs. "central" languages --- applied here not with respect to a genetic family but to a linguistic area. Wagner's views in this regard are hard to pin down, and in one respect verge on self-contradiction. His basic methodological point (regardless of margin or center) is that we should focus attention on those languages or that sprachgeographical zone where a feature is most vigorously in evidence, for here is surely the source from which the feature spread.⁸³ The view he usually espouses is that

⁸³ "Wichtig ist die Feststellung des Ursprungsherdens einer Erscheinung, der dort liegen muss, wo sie am stärksten ausgeprägt ist" (1964:245).

this "heartland" is to be found in the geographical center. Marginal languages are of course a notorious locus for archaisms, but of a particular sort: largely "isolated" features, often occurring "in rather petrified form" (1969:208; cf. 1959:121). Rather, says Wagner, in his discussion of ergativity (1978), "depending upon my experience in linguistic geography, I am inclined to believe that the explanation of a dominant feature of a particular linguistic type must be sought in a central area where it is most strongly represented and most widely spread" (41). But compare the following comment, made in the context of his areal discussion of "North European" phonetic features (1964): that "Strukturgesetze, welche in einem weitem Raume wirksam sind, an den äussersten Rändern am exzessivsten zum Ausdruck kommen" ["Structural laws obtaining over a broad area are realized in their most extreme form on the outermost margins"] (297). This would appear to contradict the 1978 statement. Whether it actually does so or not turns on an ambiguity of phrasing: does "most strongly represented" (1978) mean "represented in the majority of languages" or "realized in its most extreme ['exzessiv'?] form"? Only on the latter reading is there any contradiction: the most extreme realization of the feature would be found both in the center (1978) and on the margins (1964). And even in this case, Wagner might well have intended by "exzessiv" (1964) a hyper-normal degree of secondary development, something rather different from the vigorous preservation of an original feature apparently intended in the 1978 quote. But telling the two apart in any concrete instance would obviously not be easy.

2.4.2.2 Sprachgeographie and history

Despite Wagner's repeated plea that language be viewed in its historical matrix (e.g. 1964:284), one of the most evident features of Sprachgeographie is its dominant ahistoricity. Especially early in his career, he chooses carefully to segregate the synchronic presentation of an areal phenomenon from the diachronic recasting of that phenomenon in terms of sub/adstratal influence:

Es [geht] uns in erster Linie um geographische Zusammenhänge
..., deren individuelle historische Deutung im Grunde genommen
weniger wichtig ist und z.T. hypothetisch bleiben muss...

[We are dealing first and foremost with geographical connections,
whose individual historical explanation is at bottom
less important and must remain in part hypothetical...]

(1959:104)

One of the most surprising manifestations of this ahistoricity is Wagner's reservations vis-a-vis the term "substratum" itself. More than once he stresses that Pokornyan substratal explanation is not the best way to approach a particular typological parallel (1964:295, 1972a:67-68). It would seem that the primary goal, sufficient and worthy in itself, is the sprachgeographical analysis per se; once accomplished, we may then leave it to the substratalists to articulate a historical scenario.⁸⁴ Early in his career, this terminological aversion extends even to the pre-IE elements in the British Isles:

⁸⁴ "Damit sind wir in unserer Untersuchung an einem Punkte angelangt, wo wir das Weiterspinnen von Gedanken Substratspezialisten überlassen müssen" (1964:296).

Es wird klar, dass der Ausdruck "Substrat" in der typologischen Sprachforschung keinen Platz haben kann. [continuing in a footnote:] Er ist nur dort anwendbar, wo wir die einer Sprache zugrunde liegende Sprache konkret kennen ... Es ist unrichtig, von einem berberischen Substrat auf den brit. Inseln zu sprechen. Es kann höchstens vermutet werden, dass die Sprache oder die Sprachen, welche vor dem Indogermanischen auf den brit. Inseln gesprochen wurden, mit dem Berberischen und Semitischen typologisch verwandt gewesen sind.

[It will be clear that the term "substratum" can have no place in typological linguistics. It is only applicable in cases where we have concrete knowledge of the language underlying a given language ... It is wrong to speak of a Berber substratum on the British Isles. At most one may surmise that the pre-Indo-European language or languages spoken on the British Isles were typologically related to Berber and Semitic.] (1959:242)⁸⁵

In this case, at least, it is clear that what is being described is precisely what most linguists would call a substratum --- not specifically Berber, to be sure, but a substratal language nonetheless. In later years Wagner relaxes his attitude and uses the word now and then; thus he refers to "the substratum of Insular Celtic" (1981:63), and posits "auf den britischen Inseln ein vorkeltisches Substrat" (1977:207, also p. 204).

⁸⁵ The passage continues: "Der Begriff Substrat darf nur in der genealogischen Sprachforschung angewandt werden" --- a remark I do not understand.

When history does put in an appearance, its role is distinctively different, and quite a bit more amorphous, than in standard comparative linguistics. Historical process is little discussed; nor is chronology, whether relative or absolute. History rather becomes largely a matter of deducing and adducing various "Sprachschichten", linguistic "overlays" (presumably the detritus of successive migrations) which collectively go into the formation of a given language. Some languages are thus "archaic", in the sense of preserving structural features characteristic of an "early" Sprachschicht --- thus Wagner can speak of "the archaic languages of Western Europe, namely ... Basque and Celtic" (1981:58), or of the NW Caucasian languages as "Vertreter par excellence eines archaischen Sprachtypus" (1972a:54), where the type is called "archaic". Indeed, at times it seems that any geographical region showing great linguistic diversity is ipso facto suspect of being "archaic"; thus Wagner speaks of "'archaic' linguistic areas such as the Caucasus, New Guinea, Australia, pre-Greek Asia Minor or the pre-Roman Alps, where we find large numbers of (often unrelated or only typologically related) languages spoken by relatively small communities" (1969:229n9a).⁸⁶ The apparent assumption is that linguistic diversity is always (?) early, with regional uniformity emerging from a recent and secondary overlay. And where luxuriant linguistic diversity survives, archaic linguistic features should presumably stand a good chance of surviving too.

⁸⁶ Though the following characterization should give anyone pause: "... isolated and archaic linguistic units such as we have referred to earlier on in this paper (Amero-Indian, Chukchee, Ket, N. W. and S. Caucasian, Burushaski, Papuan, Basque, as well as Sumerian and perhaps also Hattic...)" (1976:406-7). Are half the world's languages, including the entire Western hemisphere, to be taken as "isolated and archaic"?

In actual praxis, Wagner's sprachgeographical deployment of data tends to be transhistorical, drawing freely on languages of all periods. This too is deliberate: structural features, being a reflection of "innere Sprachform", are an intrinsic part of the geographical landscape and thus should show great geographical stability over time. In his discussion of North European "Lautgeographie", Wagner is very clear about this transtemporal nature of areal linguistic analysis: "Die Sprachlandschaft ist eine Konstante, innerhalb welcher gleiche oder ähnliche Lautentwicklungen immer wieder, zu ganz verschiedenen Zeiten, stattfinden können" ["The linguistic area is a constant, within which identical or similar sound changes can recur again and again at very different periods"] (1964:251). Indeed, areal transtemporality is advertised as a strength of the method: "As linguistic areas develop the same grammatical features at different chronological stages and in successive languages, the chronological gap [some two millennia] between the periphrastic constructions of Egyptian and those of Western Europe need not surprise us"; orthodox historical linguistics, by contrast, might be inclined to reject such comparisons (1981:58-59).

This ahistoricity has the frequent consequence that a point about a language will be "proved" by adducing structural features from an earlier stage of the language's history, notably a feature implicated in one of the macro-types manifesting "innere Sprachform". Such features, it would seem, run so deep in the typological fabric of the language as to be effectively indelible. Several examples should make the point. Wagner argues, in part on historical-phonological grounds, that various preverbs and conjunct particles of Irish, e.g., negative ní-, originally incorporated a copula (1959:167): ní beir "he does not bear" was

originally (schematically) *ni-(i)s-beir. (See sec. 2.4.3 [6, ii] below.) This is then appealed to as an argument for the "verbal" nature of preverbs in Celtic. Similarly, the Berber preverbs are argued to represent frozen forms of verbs in pre-Berber; this again is taken to support the "verbal" nature of preverbs in Berber itself (1959:166). Or an example from Semitic: Wagner asserts that the "prefix conjugation" is the only "echt verbal" conjugation in Semitic (1959:200), the suffix conjugation being a nominal form (stative) (202). Though most likely correct for Proto-Semitic, this characterization holds synchronically only for Akkadian; elsewhere in Semitic, as far back as Ugaritic (14th century BC), the tense/aspect system has been totally recast, with the prefix and the suffix conjugations now on a par and straightforwardly "echt verbal".

All these cases involve the same error: the uncritical projection of a protophenomenon into some unspecified phase(s) of the language's later history. The survival of a phonological reflex of a copula, for example, in itself says nothing about the copular nature of that reflex.⁸⁷ Nor can we simply take it on faith that the Semitic language(s) which partook in a posited "eurafrikanisch" (sec. 2.4.2.3) stratum necessarily preserved the archaic, Akkadian-style tense/aspect system. That would hold true automatically only if the genesis of that stratum predated proto-Semitic; otherwise it will depend on when and how the stratum took form, and which Semitic language(s) happened to be

⁸⁷ Greene 1969 makes exactly this point in criticizing Wagner's analysis of ni. Indeed, Wagner makes the point himself: the fact that the Insular Celtic "impersonal" forms originated in IE passives is not sufficient, he says, to demonstrate that the impersonal as a syntactic category did so (1959:58).

involved. Wagner does not address these diachronic issues, and the "between-the-lines" impression one gets is that these are somehow the wrong questions; anything having to do with "innere Sprachform" should transcend temporality.

2.4.2.3 Celtic/Hamito-Semitic as a study in Sprachgeographie

Wagner's approach to the Celtic/Hamito-Semitic problem shows one striking deviation vis-a-vis Pokorny's: the pre-Insular Celtic substratum is not related to Hamito-Semitic genealogically, but only typologically. The following late summation reflects his position throughout his career,⁸⁸ and at first glance seems clear enough:

My own position in this matter, which Pokorny and myself inherited from Morris-Jones, is as follows: ... Long before the arrival of Celtic or Belgic tribes, these islands were populated by people who spoke languages or dialects which, from the point of view of E. Lewy's typology could be described as Hamito-Semitic, languages not necessarily genealogically connected with but of a similar type as Berber and Egyptian and, somewhat more distantly Hebrew and Arabic. ... When Celtic was adopted by pre-Celtic populations, the

⁸⁸ There are (at least?) two exceptions. Once Wagner ascribes Basque/Insular Celtic similarities to "a common Libyco-Berber ad- and sub-stratum" (1976:404); in 1977 he refers to "ein voridg. Substrat nordafrikanischer Herkunft" ("of North African origin") on the British Isles (204), though a few pages later it has reverted to a substratum "das typologisch mit dem nordafrikanischen Rande zusammenhängt" (207). These Pokornyan formulations are so much at variance with the rest of Wagner's oeuvre that I am inclined to take them as slips.

structure of their original language(s) began to impose itself on the language of the Celtic invaders. (1987:20, *italics mine*; see also 1959:205)

The wording is striking. Everything conspires to highlight the concept "Hamito-Semitic" (note the italicized passage); Wagner situates himself in the Pokornyan tradition and presents his argument as a variant (now typological instead of genealogical) on the Pokornyan appeal to Hamito-Semitic. Yet in fact the recast argument, if taken at face value as a typological argument, has nothing to do with Hamito-Semitic. For, if the substratal language is not actually Hamito-Semitic but only similar in type, why mention "Hamito-Semitic" at all? How does doing so illuminate the problem? Why not just characterize the pre-Celtic substratum language as "similar in type to attested Insular Celtic" (to which it surely bore a greater similarity than to Hamito-Semitic!) --- or to some hypothetical language of similar type elsewhere in the world? If Wagner nevertheless does appeal specifically to Hamito-Semitic, not just here but at every turn, it is clear that the above argument cannot be taken at face value. Plainly it is the concrete Hamito-Semitic superfamily itself, and not just the abstract Hamito-Semitic type, that is important to him.⁸⁹

There must, therefore, be something beyond pure "typologische Verwandtschaft" that links Celtic (and the pre-Celtic substratum) with Hamito-Semitic. To no surprise, for Wagner the link is areal: like all typological issues, this is a problem in Sprachgeographie. Hamito-

⁸⁹ Recall, too, the "slips" mentioned in the footnote above, where Wagner refers directly to a Hamito-Semitic substratum.

Semitic provides a concrete demonstration that there did exist in ancient times an actual language group of the "Celtic" type, and not too very far away --- obviously an important desideratum for sprachgeographical analysis. Yet such areal analysis immediately runs into an obvious problem --- which Wagner, curiously, never identifies as a problem, and to which (as we will see immediately) he ultimately can offer no good answer. The difficulty is simply that this is the type of case (sec. 2.4.2.1) where a geographical connection is not patent; the British Isles and North Africa do not form an "area" by any normal geographical criterion. A "sprachgeographisch" analysis must thus involve something more than meets the eye.

Here Wagner's point of departure is the work of his teacher Ernst Lewy --- though Wagner goes considerably beyond Lewy in what he is willing to do with the notion "area". In his major typological study of 1942, Der Bau der europäischen Sprachen, Lewy cut up the map of modern Europe with structural isoglosses that divided the territory into typological zones. One of these, the Atlantic group, comprises the "flexionsisolierend" languages of Western Europe: Basque, French, English, Scandinavian, Irish, etc., with Basque as the clearest exemplar (1942:56-57, 87)⁹⁰ --- though Lewy specifically notes (57) that Irish stands somewhat apart here. This group stands in opposition to various other typological/geographical clusters of languages, more heavily inflecting, found further to the east. Wagner takes this Atlantic group (ignoring the caveat about Irish) and expands it geographically to

⁹⁰ "Am deutlichsten tritt sie [scil. Flexionsisolierung] im Baskischen hervor, wo sie eigentlich den ganzen Sprachbau beherrscht" (56-57).

encompass North Africa as well; and it is as part of this "Eurafrican" zone that the Celtic/Hamito-Semitic similarities are to be areally understood.⁹¹ This approach would appear to dispose of the above-mentioned problem, for the languages subsumed under "Eurafrican" do form a more or less contiguous areal block.⁹² An areal analysis of this sort could thus have real explanatory value, if the relevant languages really did share striking structural parallels.

They do not, however. The set of structural parallels at issue are overwhelmingly specific to Insular Celtic and Hamito-Semitic, and not to any of the intervening languages of Western Europe --- and in particular not to Basque. Wagner himself admits in a later article that "the basic structure of Basque has little in common with these languages [viz. Celtic and Hamito-Semitic]" (1976:397). Basque does show some resemblances with Celtic; but they tend to be far fewer than, and (crucially) to overlap only partially with, the resemblances between Celtic and Hamito-Semitic. The Basque/neo-Celtic isogloss which Wagner spotlights

⁹¹ "Lewy's Western European area has an extension in Northern Africa. This would explain the syntactical similarities between the languages of Western Europe and African languages such as Egyptian and Berber" (1981:58). The same view finds expression in the title of part III.A of *Das Verbum* (152): "Das Keltische, Berberische, Baskische, Englische und Französische als Vertreter einer nordafrikanisch-westeuropäischen Sprachschicht, dargestellt am Bau des Verbums"; on p. 155 he calls these languages "eurafrikanisch". Similarly 1969:208, 1972a:55. --- It should be noted that other scholars, notably Hubschmid and Mukarovsky, have used the term "Eurafrican" to denote an old, pre-Berber and pre-Basque linguistic stratum defined in terms of vocabulary (Hubschmid 1953:92-93, 1960:39-40; cf. Craddock's discussion, 1969:42-47). Of course, such a lexically defined Eurafrican area need not be coterminous with one defined in typological terms.

⁹² The Straits of Gibraltar present no difficulty: "In terms of cultural history and cultural geography bodies of water ... have served to connect as often as to separate" (Masica 1976:9); cf. Wagner's reference to "geographisch (auch übers Meer) benachbarte Sprachen" (1972a:68).

most often, the tendency to "Flexionsisolierung" (sec. 2.4.1, cf. 1959:98-103), is a feature of neither Semitic nor Berber, though it does occur in later Egyptian and in Cushitic (cf. Wagner 1981:59). The Basque/Irish/Berber etymological equations involving -d- and/or -n- are dubious in the extreme, as argued repeatedly herein (secs. 2.3.2 [12], 2.4.3 [6 and 7]). Wagner does mention the "so-called relative forms of the verb" as a feature common to Basque, Celtic, and Berber (1976:398, cf. 1969:230n18, 1972a:37-38); likewise the existence of a polypersonal verb marking object as well as subject (1959:155). These are solid points of resemblance. But even here the comparison is less than ideal: these two features link Basque and Berber with early forms of Celtic, whereas the "Flexionsisolierung" principle links Basque with late Celtic.⁹³ All this would appear to argue that, despite certain similarities, the Atlantic group (with Basque) should if anything stand apart from Hamito-Semitic and Insular Celtic (recall Lewy's reservations about Irish), rather than embracing them as intimate typological kin.

The problem runs deeper, however. Quite apart from matters of fact, Wagner's presentation of the notion "Eurafrican" is explicitly self-contradictory, on two counts: membership in the type, and definitional characterization. The epitome of Lewy's Atlantic type was Basque (1942:56-57), and Wagner's Eurafrican type is presented as an extension of the Atlantic type.⁹⁴ Basque, then, should be for Wagner a very good

⁹³ As we will see in Chapter 6 (sec. 6.2), Basque shares additional structural isoglosses with Celtic and Hamito-Semitic, too --- but no more than do numerous other languages from all over the world. In the ranking of languages for degree of resemblance to Celtic and Hamito-Semitic, Basque is not even remotely close to the top.

⁹⁴ Wagner reiterates this at the beginning (1959:152, 155), middle (1969:208), and end (1981:58) of his career.

Eurafrican language. In the 1976 article, however, Basque is presented as having little in common with Celtic and Hamito-Semitic (397); in fact, "its impact on the Celtic and Romance languages of Western Europe is negligible compared with that of the Euraffrican substratum" (398). The wording is exceptionally revealing: "the impact of Basque" and "the impact of Euraffrican" are presented as contrasted opposites, with the logical consequence that Basque itself cannot be Euraffrican. --- The same contradiction emerges in comparing Wagner's characterization of the Euraffrican type with Lewy's description of the Atlantic type. The essential feature specifying the Euraffrican type is its "anreihend" (VO) word order: thus Wagner speaks of a "eurafrikanischen, anreihend-verbalen Sprachtypus" (1959:182, similarly 1976:402-3). However, this has nothing to do with the "Flexionsisolierung" which characterizes the Atlantic type for Lewy, and in fact directly contradicts the OV order of Basque.⁹⁵ How, then, can the Euraffrican type be an extension of the Atlantic type? One might venture the proposal that Wagner is no longer conceptualizing the "Atlantic" type in the same way as Lewy had.⁹⁶ Yet in 1981 Wagner again mentions "Flexionsisolierung" as the characteristic feature of Lewy's Atlantic group (58), with no hint that he himself sees things any differently, at the same time again presenting the Euraffrican type unproblematically as an extension of the Atlantic group (and

⁹⁵ At one point (1969:208) Wagner justifies his Euraffrican extension of Lewy's Atlantic type in yet another way, by appeal to a typological map of Nils Holmer's (Holmer, Uesson, and Smedberg 1961:31). Holmer's typology is based on prefixation vs. suffixation of various morphemes (1961:8-9); this has nothing to do with either "Flexionsisolierung" or "anreihend" (VO) word order.

⁹⁶ For example, the phenomenon of "syntaktische Anlautveränderungen", found in Insular Celtic, Basque, Berber, and Fula, is termed an "Atlantic" feature (1972a:64).

referring, in so doing, to his 1976 article!).

The contradiction seems inescapable. Eurafrican, structurally antithetical to Basque, is to be an extension of an Atlantic type whose epitome is precisely Basque. Wagner's failure to appreciate the plain paradox in his own words is very hard to understand.

None of the above considerations should be taken as invalidating the notion of a pre-Indo-European, Eurafrican VO "Sprachschicht" as a hypothetical prehistorical construct. The point is only that, contrary to Wagner, the geographical distribution of attested structural isoglosses does not in itself provide much factual support for the concept. Basque is simply a red herring: with Basque, the posited linguistic area is (barely) tenable geographically but not linguistically; without Basque, it is untenable geographically. Either way, Wagner's areal explanation is a chimera. Nor is there any reason to expect such a Sprachschicht on theoretical grounds, either. There is, after all, no reason to think that prehistoric "linguistic landscapes" in general should have been any more typologically uniform than is the case today --- and in particular, no reason a priori to insist that Europe, or any portion of it, must have been typologically uniform before the coming of the Indo-Europeans. Wagner's Eurafrican Sprachschicht is thus just as hypothetical as Pokorny's explicitly Hamito-Semitic substratum, and does not necessarily represent an advance (as implied, apparently, by Meid 1972:1197-98).⁹⁷ It may be more satisfying intellectually than a migration scenario, but it does not have a solid basis in fact.

⁹⁷ If anything, the Pokorny scenario has the advantage, inasmuch as it is specific to Insular Celtic and Hamito-Semitic.

Wagner has little to say regarding the historical process underlying his sprachgeographical analysis.⁹⁸ For the most part he makes the familiar appeal to "Sprachschichten": that the invading Indo-Europeans overlaid an earlier Western European landscape consisting of languages of a different type, a type originally flourishing in both North Africa and Western Europe but which survived only on the fringes of the European continent through substratal transmission (see e.g. 1959:102, 205; 1976:396-97; Adams 1980:48-49). Once, late in his career, he does entertain a Pokornyan migration from North Africa to Europe: "Hamitic and Proto-Hamitic speech was spread [from the Middle East] ... not only all over Northern Africa but also, if my reading of the linguistic affinities of Ancient Iberian, Basque and the substratum of Insular Celtic is correct, to Western Europe" (1981:63). This statement, however, is not aimed specifically at the genesis of a pre-Celtic British substratum, but much more broadly of a "Eurafrican" layer in Western Europe as a whole. The remark, in any event, is highly uncharacteristic.

2.4.3 Points of similarity

Das Verbum plays a role in Wagner's scholarly oeuvre analogous to that of Pokorny's article of 1926-1929: appearing relatively near the beginning of the author's career, in 1959, it constitutes the only full-length exposition of his views, with later articles being largely amplification and restatement. Unlike Pokorny's major study, however,

⁹⁸ Note that the quote which leads off this section is a historically worded statement about Celtic and Hamito-Semitic which nonetheless manages to say nothing about any historical links between the two.

it was presented in book form, and as such could evoke focused reaction in the form of book reviews (sec. 2.4.4). The summary which follows is thus based largely on Das Verbum. This is partly for the reasons just given, and partly because there is no late-in-life analogue to Pokorny's summary article of 1959,⁹⁹ but also because the book appeared in the same year as the 1959 Pokorny article; the two can thus be taken as complementary.

One general point must be mentioned in advance. Pokorny said little about the verb (beyond brief remarks in 1949 and 1964), and Wagner's work thus fills a significant gap (Pokorny says as much in his review, 1960b:141). However, Wagner's comparative verbal analysis tends very strongly to go below the surface, drawing on his own complex macro-typological views of "anreihend-verbal" languages. Much as Wagner would have disliked the characterization, such comparisons are heavily "theory-bound". Just to state the facts will thus often involve considerable preliminary ground-laying; recall the earlier discussion in sec. 2.4.¹⁰⁰

1) Celtic and Hamito-Semitic belong to the "anreihend" (ideal VO) word order type (156).

⁹⁹ Wagner's 1981 article (from a multidisciplinary conference on "The Celtic Consciousness") does provide an overview, but is basically a popularization.

¹⁰⁰ In the presentation below, page references lacking specification of year are to Das Verbum (1959) except where context clearly indicates otherwise. Hamito-Semitic comparanda focus on Berber; for a concise mini-summary see 1959:181-82, and for a summary list of features of the British verb, p. 118.

2) The Old Irish and Berber verbal complex shows identical behavior regarding the mobility of object markers --- suffixed to a simplex verb, infixes to the preverb of a complex verb (153). One could characterize these markers (though Wagner does not) as behaving like "second-position clitics".

3) In old Insular Celtic and Berber (also Semitic) --- and in Basque and French --- the verb has a "polypersonal" character, marking object as well as subject (152-56). Of these languages, the two which resemble each other most closely in this regard are Berber and Insular Celtic (155-56).

4) Basque, Celtic, and Berber have "relative forms" of the verb, specialized for use in relative clauses (1976:398; cf. 1969:230n18, 1972a:37-38).

5) Wagner notes the retreat in Old Irish, and the near-total absence in later Insular Celtic, of "semantic" preverbs, i.e., preverbs with real lexical content on the classic Indo-European model. In Old Irish such preverbs do exist, but they are much less clearly profiled than elsewhere in Indo-European due to allomorphic alternation and massive phonological fusion ("phonetisches Einschmelzen des Verbalkomplexes", 1959:122). In later Celtic these prefixes, where they survive at all, have lost their identity and become integral parts of new fused, synchronically noncomposite verb stems (168). However, Insular Celtic does have "functional" preverbs: Neg(ative), Q(uestion), various Rel(ative) markers, etc. --- just as in Berber, which also (like Hamito-Semitic generally) lacks semantic preverbs (153). Wagner invokes this parallelism to explain the non-Indo-European turn taken by Insular Celtic in

this respect: "Das Altir. verbindet also ein idg. Praefixsystem (semant. Praef.) mit einem nichtidg., das in Nordafrika zu Hause ist" ["Old Irish thus combines an IE prefix system (semantic prefixing) with a non-IE system, one which is at home in North Africa"] (166). This lack of semantic preverbs makes Insular Celtic and Hamito-Semitic "root-inflecting" in type (cf. 21).

6) Semantic preverbs are "nominal" in essence, in the sense mentioned in sec. 2.4.1 above. So it comes as no surprise that Wagner is at pains to present the functional preverbs of Modern Celtic and Berber as being "verbal" in nature: "Wir können hier von verbal-funktionellen Verbalpraefixen sprechen, im Gegensatz zu den nominal-semantischen" ["We may speak of functional-verbal verb prefixes here, in contrast to semantic-nominal"] (166). This would yield an appealing account of second-position clitics noted earlier: clitics would simply come after the initial verbal element, be it a "real" verb or a quasiverbal preverb (166, apud Pokorny 1949:236-37). Wagner argues for the verbal nature of the prefix complex from many points of view. For general methodological criticism of "preverb verbality", see sec. 2.4.1 above. Here I will respond in detail to the particular arguments Wagner presents.

i) For Berber, Wagner argues diachronically: verb prefixes are "verbal" because they are of verbal origin (166). He cites the theory of Marcy (1936:56-57), who posits pre-Berber verbal sources for several major preverbs. On Marcy's view (also picked up by Pokorny [1949:236-37]) originally only the preverb was conjugated, while the main verb was uninflected and quasi-infinitival. --- However, the preverb in fact can only conjugate for object, never subject, in synchronically attested

Berber (and only under certain conditions); while the main verb always inflects for subject, and in an inflectional paradigm demonstrably going straight back to Hamito-Semitic (see Rössler 1950), that is, to a time preceding the rise of preverbs in Berber. These factors militate against Marcy's view.

ii) For Old Irish, Wagner argues that the preverb complex in fact includes an old (impersonal) copular element --- sometimes a reflex of *is, sometimes the element *d found in forms of the copula (con-da) and recurring (in Wagner's view) in the so-called "Series B and C" infixed pronouns (Thurneysen 1946:257ff.),¹⁰¹ sometimes zero (a "virtual" copula). Regarding *is, first of all, Wagner notes (167) the lack of verb-root lenition (expected after vowels) after the Neg preverb ní- and after vowel-final semantic preverbs like do- in *do-beir "he gives" --- a phenomenon standardly explained by positing an intervening "blocking" element such as *(i)s in the early Celtic preverb complex (schematically: *do-(i)s-beir). Celticists have variously identified this element either with the copula or with an old 3-sg pronominal element. For Wagner the copular analysis is plainly correct; the pronominal view "ist aus der Luft gegriffen" (168).¹⁰²

iii) Regarding Irish *d, Wagner (170) proposes that a verbal form

¹⁰¹ By 1967, Wagner's diachronic identification of infixed-pronominal *d with copular *d is less absolute: he now speaks of the two as being "mindestens synchronisch gesehen dasselbe Element", and continues: "Was dieses d-Element historisch einmal gewesen ist, spielt dabei eine geringe Rolle" (1967:304-5).

¹⁰² Two decades later, Wagner continues to maintain this position, but admits that the pronominal view is "lautlich ... sicher befriedigender" (1977:210n21).

with a Series-B infix, such as -don- in

for-don-cain "he teaches us" (normally so glossed)

over-us-sings ,

is better taken as

for-d(o)-n-cain

over-is-us-sings.

He rejects (171-72) Thurneysen's view of a division into two essentially homophonous infix series B and C, for these series would in fact differ only with regard to mutations of the leading element -d-. Occam's Razor argues against this division, for these mutations have a ready explanation: the -d- quite naturally occurs unmutated in main clauses, lenited in leniting relative clauses, and nasalized in nasalizing relative clauses. This is exactly the behavior to be expected, in a non-infix compound verb, of the verb root itself (e.g. for-cain, where it is the c which would occur plain or mutated) --- thus demonstrating, in Wagner's view, the inherently verbal nature of -d-. But the conclusion is a non sequitur. The "unification" argument for series B and C is indeed attractive (Gagnepain endorses it [1961:325]), but nothing in it hinges on a copular analysis of -d-. The standard formulation (following Thurneysen) of the relative-clause mutation rule, i.e. in strictly linear and non-categorical terms, takes the mutations as simply affecting "the following [scil. post-preverb] initial" (1946:314), regardless of the identity of this initial as verb root or infixed pronoun. There seems no reason to recast this formulation.

iv) As we have seen (Pokorny, sec. 2.3.2 [12]), the Irish element *d has a correspondent in Berber with a variety of uses. Wagner focuses in

particular on the Berber "d of predication" appearing before predicate nouns or adjectives; this he calls an impersonal copula (173) or "partikelartige Copula" (177). Wagner here cites the great Berber authority Basset (1952:38), yet without noting Basset's arguments against a copular analysis of predicative d (inter alia, d can cooccur with a true copular verb). Wagner further suggests a striking similarity between Irish and Berber as regards the role of this impersonal copula in clefting constructions ("mise en relief": "(It) is John who..."): "bis in Einzelheiten ist die ir. Konstruktion mit der berber. identisch (impersonelle Copula ir. is, berber. d; proklitische Relativpartikel; finite Relativform)" (173). He acknowledges the criticism that clefting (Copula + Noun/Pron + Rel + V) is common crosslinguistically, but nonetheless sees it as a distinctively "anreihend" strategy for emphasis and thus especially "well anchored" in Berber and Insular Celtic (and French!) --- as opposed to (say) German, where strong phonetic stress is the favored technique for emphasis.

v) In both Berber and Celtic, Wagner proposes to connect such "partikelartige" copulas to deictic elements; this would explain in a natural way the copula's impersonal nature (176). In Berber, the "copular" d (= predicative particle) is homophonous (thus "identisch") with the infixing/prefixing proximal deictic d "here" (175). In Celtic, Wagner does not deal with d, but points to the derivation of the suppletive Welsh y mae "it/he is" from yma "here"; of Welsh llyma "voici" from a sequence syll yma "see here"; and of Old Irish fil, the suppletive stem of the substantive verb "be", from *wel- "see" (176).¹⁰³

¹⁰³ Note that the last item (fil) involves no deictic at all, even though the rhetorical context is one of establishing a link between

vi) Wagner reinforces the Irish-Berber parallel by pointing out (177-78) the existence in both languages of a dichotomy between copula (impersonal) and substantival verb --- clearly true for Irish, but in Berber dependent on the analysis of d as copular. Spanish estar/ser is mentioned, as is the Semitic etymology of the Arabic copula kāna from "to stand".

7) Like Pokorny, Wagner proposes a Berber/Irish/Basque link not only for the element d but also for n (178ff.). For d see [6] above, notably with regard to its "copular" nature in the two languages, which for Wagner lies at the heart of the parallelism (recall also sec. 2.3.2 [12]). The discussion of n is some of the most inscrutable in Wagner's writing. For Wagner (178) the Berber proximal deictic n is identical with the Berber genitival n, and the latter in turn with the suffixed -n occurring in the "participial" verb form characteristic of subject-relative clauses. In support of the latter equation, he invokes "die engen Beziehungen zwischen Relativsatz und Genetiv im Hamito-Semitischen", which are "allzu bekannt" ["the close ties between relative clause and genitive in Hamito-Semitic ... which are all too well-known"]. This refers to the Semitic/Egyptian phenomenon whereby the selfsame particle transparently functions as a preposed marker of both copula and deictic. This is "associative" argumentation at its most extreme: the topic flicks from copula-as-deictic to copula-as-"see"+deictic to copula-as-"see". Wagner immediately backs up Irish fil with a parallel: the Egyptian m-k, a "pictorial" particle rendered as "behold + <2-sg>" (Gardiner 1957:178). For Wagner (though not for Gardiner) this is a copula; further, Gardiner's tentative etymological derivation ("possibly an obsolete imperative meaning 'behold'") has metamorphosed into something much firmer: "Gardiner analysiert ... richtig als 'sieh-du mich!'" French voilà is also mentioned (here a deictic element does occur), as are links with Fula (progressive marker d, adverb do "here"). All this goes by in less than a page (176).

genitive and relative embeddings ("of" = "which"). The Berber phenomenon, however, is not structurally parallel: relative -n is usually a suffix, and hardly a "relative particle" in the transparent Semitic/Egyptian sense.

Grant Wagner his argument; we then have for Berber the equation

$$\underline{n}: \text{deictic} = \text{genitive} = \text{relative}.$$

Recall, too, that the element d satisfies the equation

$$\underline{d}: \text{deictic} = \text{copula}.$$

Appealing to the shared deictic nature of Berber n and d, Wagner proposes that "in diesen typisch anreihenden Sprachen zwischen Relativzeichen und Copula eine enge bedeutungsmässige Beziehung besteht" ["there exists a close semantic link between relative marker and copula in these typically 'anreihend' languages"] (179). He does not amplify on this leap of logic; note that, based on the data presented, neither element (d, n) encompasses both relative and copula function.¹⁰⁴ Note further that this is a purely Berber-based argument; yet Wagner presents his conclusion as applying to "anreihenden Sprachen" (plural). In Old Irish a similar link is then proposed (though not involving the same etymon) between the "relative ending" -es and the formally identical copula. Likewise in Modern Irish: the relative particle a (schwa) is actually a form of the copula, says Wagner, because it has a past tense form ar (181). Irish has no genitival n; nonetheless, Wagner does propose a hodgepodge of fuzzy links (termed "deutlich") between genitive

¹⁰⁴ Here deictic function would seem to be the sole point of commonality. In fact, some Berber dialects do have a relative particle in d (see Sadiqi 1986:27ff.); but Wagner, as far as I can tell, nowhere mentions this fact.

and relative in Irish (179-80). It falls out, too, that Berber copular/deictic d can be compared to the Ugaritic (actually pan-Semitic) genitive/relative marker d.

This is Wagner's "associative" style of argumentation at its most exuberant.

8) In prepositional relative clauses (the type "the man with whom you went"), Berber and Old Irish both form the construction by fronting the bare preposition: schematically,

Berber: man (REL) with [you went ____]

Irish: man with-REL [you went ____]

In neither case is there a resumptive pronoun; further, the REL element is an invariant relative particle, not a relative pronoun. In Berber, says Wagner, "wirkt ... die betreffende Praeposition als Relativum" ["the preposition acts as relative marker"]. In Old Irish, where Prep+REL occurs preverbally in relative function, the REL marker (identical in form to a frozen neuter definite article) is actually "kein eigentliches Relativelement" (180), evidently implying (though Wagner does not say so) that the "real" relative element is again the preposition. Regardless of the analysis, however, the type is clearly very different from Indo-European and represents a striking commonality. Elsewhere in Celtic and Hamito-Semitic, a quite different oblique relative construction occurs (Copying), equally alien to Indo-European.

9) Wagner's aspectual analysis of the verb in modern Insular Celtic, Berber, and Semitic presents the unmarked finite verb as fundamentally "aktionell" or punctive; durative categories are seen as marked and

derived ("habitual" seems to go both ways). This theme is developed at great length in Das Verbum for Celtic (Part I) and Hamito-Semitic (Parts III.A and B); only a capsule sketch can be given here.¹⁰⁵

For Insular Celtic, Modern Irish must suffice by way of illustration.¹⁰⁶ Here the primary verbal tense forms, "die eigentlichen zéro-Formen" (22), are the synthetic future and preterite, for only they simply lay out the verbal action with no aspectual nuancing --- that is, punctively, since "die Aktion an sich" is always punctive (22). The synthetic present, which serves as a habitual or timeless present, is deemed non-primary because aspectually marked (30).¹⁰⁷ Duratives, by contrast, are formed periphrastically ("he is at VERB-ing") and as such are secondary.

In Berber the basic form is the aorist, an inherently punctive expression of the action per se (184). This is used with preverb a(d) for voluntative, future, and complement-clause functions, and without preverb for continuative ("and-then") function (184). Some verbs also have a distinct perfect (= preterite) form for narrative past time (185); for most, however, the aorist fulfills this function too.

¹⁰⁵ Wagner presents brief contrastive summaries of aspect on pp. 188, 204-5. On the Hamito-Semitic verb see Rössler 1950, which Wagner frequently draws on, and Rössler's favorable review (1960) of Das Verbum. A critique of Wagner's aspectual analyses would make a book in itself.

¹⁰⁶ Wagner presents an overall summary of "British" verbal features on p. 118; charts of verbal tense-aspect appear on pp. 59-60 (Irish), 64 (Welsh), 86 (Scots Gaelic). For diachronic developments see especially discussion on pp. 66, 84, 189.

¹⁰⁷ The synthetic present is "marked" only semantically, not morphologically; elsewhere Wagner characterizes the Irish habitual forms as "nur bedeutungsmässig charakterisiert", in contrast to their Berber counterparts (below), which are in addition marked morphologically (188).

Opposed to the punctive macro-category "aorist/perfect" is a durative form: the "habitual" (termed "aorist intensive" by Berberists), a morphologically marked category formed by various root modifications (affixes and internal changes, including gemination) (187).

Complicating the analysis of tense/aspect in Semitic is the sharp split between Akkadian and West Semitic (with Ethiopic a special case):

	Akkadian	West Semitic	Ethiopic
Form:			
i-prus	Aorist	Imperfect	Subjunctive
i-parras	Present	---	Imperfect
pars-āku	Stative	Perfect	Perfect

Wagner, following Rössler (1950) and probably most other Semitists, sees the original Semitic system as fundamentally that of Akkadian. The basic verbal form is the prefix conjugation (specifically the aorist i-prus), "die einzige echt verbale Konjugation" ["the only truly verbal conjugation"] (200, cf. 204). Drawing on its uses in both Akkadian (past narrative, voluntative) and West Semitic (voluntative, future; also continuative in Hebrew), Wagner presents this form as the formal, functional, and diachronic correspondent of the Berber aorist. By contrast, "durative" categories (broadly conceived) are derived and marked: either nominalized forms, as in the Akkadian stative, or "morphologisch und bedeutungsmässig charakterisierte, iterativ-intensiv-aktuelle Aktionsarten" (204), as in the Akkadian present and the Berber habitual

(both featuring gemination), which are equated diachronically (187-88, 203, apud Rössler 1950:467, 469).¹⁰⁸ The West Semitic system, wherein the "punctive" aorist (iprus) has metamorphosed into a more durative imperfect, is secondary.

Wagner clearly intends a close aspectual parallel here between Insular Celtic and Hamito-Semitic. But later in the book, in his sprachgeographical treatment of Indo-European as originally similar in type to Hamito-Semitic, he weakens his case by presenting a similar aspectual analysis for Indo-European as a whole: "[IE] Durative Verbalformen sind wie im Hamito-Semit., im Gegensatz zum Imperativ-Aorist, abgeleitet-charakterisiert" ["In contrast to the imperative-aorist, [IE] durative verb forms are, as in Hamito-Semitic, derived and marked"] (250). Clearly, the specialness of a Celtic/Hamito-Semitic aspectual parallel is undermined if essentially the same thing holds for Indo-European.

10) Wagner devotes a large section of Das Verbum (III.D) to sentence-level syntactic comparison of Celtic and Hamito-Semitic. The various points he lists (some not involving any actual "comparison" and hence omitted here) are all subsumed under the typological macro-principle of "durchbrochener Parallelismus" ["anti-parallelism"],¹⁰⁹ whereby two units that in principle ought to be structurally parallel in fact are not so. Such anti-parallelism, he suggests, is in fact a fundamental

¹⁰⁸ The geminate of the Akkadian present is synchronically quite distinct from that characterizing D-stem (intensive) verbs, where the geminate runs through the entire paradigm.

¹⁰⁹ Gagnepain translates "anacoluthie" (1961:326); Wagner elsewhere (218) uses the paraphrase "entparallelisierend" ["de-parallelizing"]. Some of the suggested manifestations of this principle can be quite baffling.

characteristic of these "anreihend" languages (207).

One such feature is the phenomenon whereby a finite verb is "continued" by a verbal noun (207-8); schematically, "he entered and to-speak". The initial verb provides tense, aspect, and person information for the following non-finite form. Drawing on the Lewy/Finck typological nomenclature, Wagner calls this a "group-inflecting" principle. It is common throughout Celtic, especially Welsh, and is found in certain Semitic languages as well (Hebrew, Ugaritic, [also Phoenician]). It also occurs in Old Norse.¹¹⁰ --- However, this feature (called "clause chaining" in today's terminology) is certainly not specific to "anreihend" languages. The mirror-image phenomenon is exceptionally common in OV languages ("to-enter and he spoke"), with the nonfinal verbs categorially impoverished vis-a-vis the final verb.¹¹¹

11) A second trait representing "durchbrochener Parallelismus" (208-11) is the existence of a special "continuative" finite verb form (an "and-then" tense), a feature found in many "anreihend" languages: Bantu (e.g. the Swahili -ka- tense), Biblical Hebrew (the tense inversion occurring after waw "and"),¹¹² Fula, Berber (the bare continuative aorist).

¹¹⁰ Old Norse shows other typological affinities to languages of Britain, as Wagner notes passim (see below); his brief attempt at explanation appeals to Viking expansion into northwest Europe (122).

¹¹¹ Wagner correctly mentions Basque in this connection (208), but presents the phenomenon as involving a finite initial verb; this contradicts the description in Saltarelli's grammar of Basque (1988:246-47), where only the final verb is said to take the person/number Aux.

¹¹² In Hebrew narrative, waw + Imperfect continues an action begun in the Perfect (typically past), and waw + Perfect continues an action begun in the Imperfect (typically future). In both, the morphological verb-form accompanying waw differs somewhat from the ordinary Perfect or Imperfect, thus conferring on the construction its own clear morphological profile.

Wagner proposes to couple this phenomenon with the common Old Irish switch between narrative preterite and "historical present", with the latter cast as a continuative tense. (Again, he remarks, also in Old Norse.) Such continuatives are essentially durative (211), says Wagner, for the continued action is conceptualized as an accompanying action (Begleithandlung).¹¹³ --- Wagner does not, however, make it clear to what extent the Irish "historical present" is truly grammaticized as a "continuative tense-form". Unlike the very clearly profiled Bantu and Hebrew phenomena, the Irish involves no special morphological form. And the sequence "narr.pret + hist.pres" is not the only one to appear; "hist.pres + narr.pret" can occur as well, if the second verb is preceded by co "until, so that, and" (216). Historical presents, indeed, would seem to be rather common crosslinguistically, typically with an added nuance of vividness which is absent with canonical continuative tenses.

12) In Irish and Welsh, the construction "and + non-finite verb" is a very common way of expressing concomitant circumstance --- the type "I met Barra, and he walking on the sea" (212-14). That this construction demands a non-finite verbal form is quite natural for Wagner: the conjunction "and" itself (like the functional preverbs, sec. 2.4.3 [6]) fulfills the function of verbal element (213). Again the pattern exists in Old Norse, and (surprisingly) even in Old English. Semitic, and

¹¹³ "Die auf eine charakterisierte punktuelle Verbalform folgende Verbalhandlung wird also nicht als selbständige Handlung, sondern mehr als Begleithandlung aufgefasst --- und ist deshalb an sich eher durativ" (211). The logic makes little sense. A continued action may perhaps be notionally backgrounded to the initial action, but it can easily be punctive, and is explicitly not an "accompanying" action (Begleithandlung) but an ensuing action ("He came in and sat down").

especially Arabic, offers a close parallel (the hāl construction); Wagner also cites a Berber example. --- Wagner fails to note, however, that in the Arabic hāl the "and" element can also be followed by a finite verb. Nor is it explained why this construction counts as "durchbrochener Parallelismus"; the two verbs are not conceptually parallel, for one expresses concomitant circumstance.

13) In modern Insular Celtic, West Semitic, and Berber, says Wagner (218), there is a rule that a conjoined (thus plural) subject nonetheless takes its verb in the singular; the rule is not yet in force in Old Irish, but is asserted to exist in Old Norse. Wagner explains this anti-parallel treatment of conjoined subjects cryptically: "Die Regel ergibt sich aus dem entparallelisierenden Charakter des 'Begriffes' und, der in diesen Sprachen mit dem Begriff mit verwandt ist" ["The rule results from the de-parallelizing nature of the 'concept' and, which in these languages is related to the concept with"]. The (implicit) logic is sound: if the syntagm "X and Y" were really "X with Y", as in Berber, then the only noun in direct construction with the verb would be X, obviating any need for plural concord. The anti-parallelism would apply to the non-parallel syntactic status of the two notionally parallel conjuncts. But the facts are not as Wagner presents them, on three counts. First, a link between "and" and "with", though palpable in Berber (the preposition "with" d is often translatable as "and") and Welsh (ac in the same two functions) (212),¹¹⁴ is nonexistent in Semitic and hardly transparent in Irish (Wagner links Irish ogus "and" with the homophonous

¹¹⁴ Berber d can only "conjoin" two nouns, never verbs or clauses, and hence I see no reason to regard it as anything other than a preposition. Welsh ac is under no such constraint.

word for "near", itself connected with Welsh ac "and, with" [212]).

Second, in Semitic the stated concord pattern is a rule only in Arabic (in Hebrew it is merely a possible variant). Third, the specific category of conjoined subjects is relevant to concord only in Berber, and even there only by the trick of recasting the preposition "with" as a true "and"-word; in Arabic and Insular Celtic, singular concord applies with any kind of plural subject whatever. Still, there is a syntactic commonality lurking here --- the general occurrence of a singular verb before plural subject in Insular Celtic and Arabic (not Berber).¹¹⁵

14) Wagner compares (219) the constellation of "archaic" verb-final phenomena in Old Irish (Bergin's Law, tmesis) to a similar verb-final construction in Ugaritic involving "emphatic k" (Gordon 1965:76, 119):

1-	ktp	^C nt	k -	^V tst	- h	"she sets him on the
	on-shoulder	Anat	PTCL-she.sets-him			shoulder of Anat"

Wagner (contra Gordon) sees this k as identical to the ordinary subordinating conjunction k, and hence takes the construction as a cleft (mise en relief), "(it is) on the shoulder of Anat that she sets him", parallel to his construal of the Irish OV constructions as clefts (1967, 1977). The same emphatic k construction is found in Hebrew (Gordon 1965:76, cf. Gesenius 1910:471). However, none of these authors discusses the frequency of this Ugaritic/Hebrew OV construction; my impression is that it is as uncommon as its Irish counterpart. A substratal or areal explanation is seldom persuasive if the phenomenon in

¹¹⁵ Note that in Egyptian the verb is totally devoid of all person/number marking (not even 3-sg) when any explicit full-NP subject is present.

question is rare in both languages. Furthermore, the Semitic construction would seem to be manifesting, not so much "OV syntax" per se, as rather the general flexibility in clause-level word order that is characteristic of most Semitic languages (Gensler 1989), notably in poetic style (as above). --- Because clefting breaks apart the conceptual unity of the sentence (219) ("ein sachlich einheitlicher Satz [wird] auseinandergerissen (segmentiert)"), Wagner sees this construction, too, as depending on the principle of "durchbrochener Parallelismus". The reasoning seems highly strained.

15) It is characteristic of Irish for adverbs to be used predicatively --- constructions of the type "(it) is likely that...", or the very common is amhlaidh... ["(it) is so, that..."]. Wagner sees in this a manifestation of the "verbal" character of Irish (222). Drawing a parallel, he says that in Berber "eine bedeutende Anzahl" ["a significant number"] of adverbs are frozen verb-forms, citing Basset 1952:41-42 (e.g. "few" < "be-few"); by contrast, "nominal" adverbs are deemed alien to Berber ("ebensowenig zu Hause wie nominale Adjektive" [223]). But Basset actually says nothing about "a significant number" of such deverbal adverbs, indicating only that, as fossils, they are hard to detect (hence hard to enumerate); by contrast, he does expressly devote a paragraph to the topic of denominal adverbs (e.g. "this moment" > "now"). Nothing in Basset's discussion indicates that deverbal adverbs are any more firmly entrenched in Berber than denominal adverbs. --- Quite apart from misrepresenting Basset, Wagner's presentation of the two languages as sharing a feature "verb-like adverbs" makes little sense logically: a posited diachronic origin for adverbs in stative verbs (Berber) is something totally different from a synchronically transparent construction

"Copula + Adv" (Irish). (See sec. 2.4.2.2 for Wagner's transhistorical (mis)use of data in linguistic argumentation.)

16) Insular Celtic languages, and Irish in particular, barely have words for "yes" and "no"; instead, yes-no questions are typically answered by repeating the verb. Similar phenomena are cited for Arabic and Hebrew. But Wagner does not mention that these Semitic languages, even in their early Koranic and Biblical forms, do have ordinary words for "yes" (Arabic) and "no" (Arabic, Hebrew) that can comprise full answers to questions (though I do not know the relative frequency of answer-types). Wagner makes a similar surmise for Berber, based on his reading of texts (223). Detailed examination of the various answering strategies in Welsh and Irish follows (223-31), but with little further comparison to Hamito-Semitic.

17) Wagner draws a phonological parallel between Berber and Celtic "consonantal weakening" in intervocalic position (sonorization, spirantization), a phenomenon he considers "typisch westeuropäisch" (228); the issue recurs in a later article (1964:297-98). Sommerfelt in rebuttal points out the same phenomenon in Northern Europe (a linguistic "Sprachlandschaft" which Wagner sets off against the "eurafrikanisch" zone), continuing: "Indeed, the development from stop to fricative is such a common occurrence in the most various linguistic areas of the world that it has little value as a typological criterion" (1968:258).

18) The Insular Celtic verbal noun (VN) represents a significant departure from Indo-European, and Wagner (232) rightly criticizes Dillon (1955:111) for accepting Zimmer's view that it is simply an archaism. For despite undoubtedness formal archaicism --- reminiscent of the

Vedic infinitive in its transparently substantival nature (232) and its multiplicity of forms (235) --- the VN represents a major innovation by virtue of having functionally absorbed the active participle, which vanished as a verbal category in Celtic prehistory. The twin use of English -ing, a verbal-noun form which took over the sphere of the old present participle, is considered by Wagner as part of the same syndrome.¹¹⁶ It is curious that the consistently genitival rection of the direct object, which Wagner does mention (233), is not presented as an innovative aspect of the VN.

Berber offers nothing similar. The Arabic VN, on the other hand, provides a good parallel, especially in light of its formal multiplicity, analogous to the situation in Irish (235). Both in Arabic and in Celtic, the VN is independent of tense and voice. In Egyptian, too, the parallels are very close. Wagner notes the use of the VN in place of a finite verb in Egyptian narrative "to announce incidents of outstanding importance" (Gardiner 1957:230), a usage closely paralleled in the Irish annals (235); in both languages, further, the subject of the VN is introduced by a preposition, while the object occurs as a genitive (236). On the other hand, active participles are alive and well in both Semitic and Egyptian, weakening the similarity to Celtic.

19) Irish, Basque, and Berber show a parallel incorporation of various conjunctions into the verbal complex (1972a:3-13 on Irish and Basque;

¹¹⁶ Significantly, Insular Celtic insists on a preposition ("he is at going", etc.), like earlier English "a-going" (< "on going"). This suggests, contra Wagner, that the Celtic VN really is a verbal noun pure and simple, with nothing "participial" about it; it is the construction "Prep + VN", not the category "VN" itself, that has supplanted the older participle.

1976:404-6 adds Berber). The Irish preposition co n- "to", when used as a conjunction "so that", has the status of a conjunct particle (Thurneysen 1946:28) and as such is an integral part of the preverb complex. Basque too is said to show the same phenomenon, only in mirror-image: the "completive" verb ending -la, here presented as identical to the allative -ra/la¹¹⁷ and as such analogous to the Irish, is again an integral part of the verbal form. Finally, the Berber preverbal particle a(d)-, indicating consecutive or completive action (1976:404) and thus functionally parallel to Irish co n-, is similarly part of the verbal complex. Common to all these cases is the circumstance that "conjunctive elements are ... integral parts of the verbal form and not independent elements of the sentence" (1976:406). Basque and Irish have the further parallel that the conjunctive element is the same as an adposition; this does not hold for Berber.¹¹⁸

20) Both Celtic and Berber have "Anlautveränderung" (1959:160-61), i.e., word-initial changes which are exploited for syntactic purposes (though the Berber phenomenon does not involve consonant mutation, but vowel or glide alternations in initial syllables); the Celtic phenomenon is "ein absolutes Unikum" within Europe and within Indo-European (121). Elsewhere (1976:402-4) Wagner casts Anlautveränderung in a different role, as the definitional criterion for a special "Western Atlantic" subtype of his "African-Semitic" (Eurafrican) group. This subtype, basically North African, allegedly "has affected Insular Celtic" (403); Berber,

¹¹⁷ The usual form of the allative is -ra, with -la a phonologically conditioned allomorph (1972a:12).

¹¹⁸ See Rössler 1960:146, objecting to the link between preverb a(d)- and deictic d in Berber proposed in Wagner 1959:165.

however, goes unmentioned.

The presence of "Anlautveränderung" is indeed a remarkable similarity between Insular Celtic and Berber. Wagner's discussion, however, only touches the surface. First, he does not distinguish between the kind of word-initial change whose functional scope is restricted to a single grammatical category (e.g. person, as in Tubu), and the kind which pervades a language's grammar in a multiplicity of uses. The former is less rare crosslinguistically, showing up in far-flung languages --- in Africa (Tubu [person] and Fula [number], per Wagner 1976:403), in Southeast Asia (e.g. Burmese [causativity]), in Oceania (Lynch 1975 [realis/irrealis]), allegedly in Basque (marginally, per Wagner [diminutive]).¹¹⁹ The latter, multicategorical phenomenon, on the other hand, is extremely rare; outside of Insular Celtic and Berber, it occurs in my sample only in the Siberian language Gilyak (see Appendix 2). --- Further, the Celtic and Berber phenomena are remarkable not only for similarities but also for differences. As I have shown elsewhere (Gensler 1992), Welsh and Berber characteristically pick inverse environments to mark with the "initial-changed" form. Wagner, on the other hand, tries to portray Celtic and Berber as being controlled by a common macro-principle: "Sowohl im Kelt. als auch im Berber. verändern grammatisch regierende Wörter den Anlaut des nachfolgenden, abhängigen Wortes" ["In both Celtic and Berber, grammatically governing words alter the beginning of the following, dependent word"] (160). This is indeed true for the objects of (most) prepositions, as Wagner asserts; but he

¹¹⁹ Contra Wagner, the Basque alternation is not specific to word-initial consonants, but may occur medially as well; see Basque data summary in Appendix 2.

makes no mention of sentential subject and object, which are marked inversely in Welsh (object mutation) and in Berber (subject "mutation"), thus contravening his generalization; and his appeal to examples like Irish mil bheach ("bees' honey"), where an embedded genitive beach ("bees") mutates under the control of its preceding governing noun, handily ignores the fact that in Welsh non-mutation of such genitives is the rule, and indeed in Irish itself the mutation occurs only under particular circumstances (see Christian Brothers 1980:15). Wagner claims, further, that initial change affects verbs as well as nouns, not only in Celtic (uncontroversially) but also in Berber; as evidence he cites the "initial" u/a alternation of Berber

uf-iʃ "I found" (Preterite) vs. af "find!" (Imperative).

But this is actually an internal vowel alternation (Rössler 1960:146), applied here to the fortuitously V-initial root af (further, the segmentation should be ufi-ʃ).

21) In Insular Celtic, the word as such has relatively little autonomy, says Wagner --- it is "ein typisches Satzwort ..., d.h. es ist lautlich und grammatisch maximal vom Satzzusammenhang abhängig" (emphasis mine) ["a typical 'sentence-word' ... that is, it is phonically and grammatically maximally dependent on the sentence context"] (81, cf. 175). In support, Wagner cites the difficulty of eliciting pausal forms of words in Irish (175), given that the "word" really is a cluster of context-dependent variants; also the lenition of the object in Welsh, which has the effect of welding words into a word-group (81). The verb in Berber (and in Bantu) is likewise characterized as a "Satzwort", Wagner says,

because of its "group-inflecting" nature (205).¹²⁰ There is an intuitive plausibility in regarding "group-inflection" as something that binds a word-group together closely (as a "Satzwort"); but it is not clear how or why Berber should count as "group-inflecting". Elsewhere Wagner argues that the Berber verb counts as a "Satzwort" because the occurrence of a given verbal form is strongly dependent on syntactic position and "Satzzusammenhang";¹²¹ I do not understand this argument.

The notion "Satzwort" is a long-established part of the European typological tradition (cf. sec. 2.3.2 [3]), and to Wagner it clearly makes intuitive sense. He presents the concept without defining it (175), then repeatedly invokes this nondefinition as a "definition" (thus on pp. 81, 183, 205). In actual usage, "Satzwort" appears to be a nebulous macro-term covering any notion of interword coherence whatever, whether phonological, morphological, or syntactic --- whence the variety of ad hoc arguments adduced above.

2.4.4 Reactions

Wagner's Das Verbum triggered a storm of critical response, ranging from rave reviews (Ó Cléirigh 1959, Pokorny 1960b, Rössler 1960) to stern disapproval on the part of "mainstream" historical linguists (notably Jackson 1962; to a considerable degree also Schmid 1960, Schmidt

¹²⁰ "Group-inflection", another term from the Lewy/Finck lexicon, is illustrated by such English constructions as "the king of France's hat", where a whole word-group gets an inflectional mark as if it were a single word.

¹²¹ "Das Auftreten einer bestimmten Form hängt stark von ihrer syntaktischen Stellung ab: das Verbum ist im Berber. ein Satzwort" (183).

1963) to mixed reactions (Gagnepain 1961, Sommerfelt 1962; also Meid 1963:50-52).¹²² Little is said about the linguistic facts assembled in the book except to marvel at their scope and to accept them (with minor corrections) as accurate. The critical focus, rather, is on method and interpretation.

The most consistently positive consensus among critics is their praise for Wagner's language-specific (and/or family-specific) analyses of verb systems and verbal tense-aspect, especially in Celtic (Jackson 1962:230, Schmid 1960:312, Ó Cléirigh 1959:125; also Greene 1966a:130) but also in Berber (Rössler 1960). "Never have the affinities between the verbal systems of Irish and English, on the one hand, and those of Welsh and Scottish Gaelic, on the other, been so clearly brought out," says Greene. Even Jackson can admit that "the account of these various [Insular Celtic] verbal systems will constitute for some readers the chief value of the book, especially the comparative treatments of Irish and Welsh," and approve of Wagner's "clear formulation of the problems involved" (230); similarly Schmid (312). Significantly, Sommerfelt commends Wagner for "having taken all the main types of the known languages of these areas [scil. the British Isles] into consideration. Hitherto one had mainly pointed to details" (1962:230).

Sommerfelt's remark points out another aspect of Wagner's work which has impressed critics favorably: the breadth and innovativeness of

¹²² Gutenbrunner's favorable review (1961/62) is too brief to warrant more than a mention. Meid 1963 reacts only to Wagner's discussion (219-22) of the Old Irish absolute/conjunct dichotomy. Two articles (in English) responding in detail to Wagner's later work are Oftedal 1968 (on Wagner 1964) and Greene 1977 (largely on Wagner 1967), both sympathetic yet critical.

his thinking, the ability to transcend minutiae and philological parochialism in favor of the big picture.¹²³ The very fact of concentrating on syntax and function was itself a significant innovation (Ó Cléirigh 1959:124). But to analyze the functioning of entire syntactic systems, compare them both within and outside of Celtic, and situate the ensemble in areal terms --- all this was radically new.

Not surprisingly, it is when Wagner's geographical focus shifts from micro to macro that the critical response becomes most negative. The areal treatment of the Celtic verb is solid and valuable (for Jackson's ambivalence, see below), and Wagner's analysis of the English verb as the product of Sprachmischung with an Insular Celtic substratum¹²⁴ also comes in for praise (Schmidt 1963:351). But Wagner's sprachgeographical treatment of Indo-European prehistory is hypothetical and unprovable: Sommerfelt (229), Schmidt (351), Schmid (316). It is illegitimate, Schmid emphasizes (315), to retain the traditional view of an "Indo-European" based on Greek, Latin, and Sanskrit, with deviations (Celtic, Hittite) automatically ascribed to areal factors. And the weaknesses of Wagner's treatment of Hamito-Semitic links and the Eurafrian "Sprachschicht", especially as regards underlying method and theory, are all too obvious.

¹²³ Wagner displays "einen guten Blick für das Allgemeine," says Schmid in an otherwise critical review (1960:316). "Il a le mérite --- qui n'est pas mince --- d'oser sortir, enfin, les études celtiques du souci exclusif de la philologie et de l'asservissement au détail" (Gagnepain 1961:326). And Rössler's detailed remarks on Wagner's analysis of the Berber verb emerge "aus der Freude ... dass endlich über diese Fragen ein Gespräch eröffnet ist" (1960:142).

¹²⁴ On this see especially Preusler 1956.

The most telling theoretical attack is concentrated on Wagner's uncritical stance toward "Sprachgeographie".¹²⁵ As Schmid puts it (315), Wagner's dictum that "Der Typus einer Sprache ist durch deren geographische Lage bestimmt" (1959:102) is just as one-sided and dogmatic as the claim that a human being's character is determined by his environment. Totally unconnected languages can develop in similar ways simply owing to the limited possible variety of language structure (Gagnepain 311, Schmidt 348-49); change may also be driven by language-internal, structural factors (Schmidt 348). Further, any valid attempt at areal explanation of typological commonalities must also look beyond the languages being compared; if the phenomenon recurs in other languages, its contribution to an areal argument is nil ("so wird die Tatsache der Übereinstimmung in den benachbarten oder verglichenen Sprachen als Kriterium entwertet" [Schmidt 350, cf. Schmid 314]). For example, Wagner's areal approach to the co-presence of Neg verb-prefixes in both Old Irish and Berber is vitiated by the existence of the same phenomenon in Lithuanian and Armenian (Schmidt 350); infixed pronouns, another striking Irish-Berber isogloss, are also found in Lithuanian (Schmid 314-15); and indeed, of Wagner's eight characteristic features of the British verb (1959:118), four recur in Hindi (Schmid 315). Finally, the choice of different traits as features of comparison would inevitably lead to different groupings of languages (Gagnepain 311).

¹²⁵ Schmid refers pungently to the "noch unkultivierten Sumpfboden der Sprachgeographie" (316). On the other hand, Gagnepain (like Pokorny and Ó Cléirigh) speaks with overall approval of the "sprachgeographisch" approach: Wagner has produced "une sorte de grammaire comparée dans l'espace, et non plus dans le temps"; Gaelic and Brythonic appear as "pôles d'attraction entre lesquels s'ordonne ... la variété des aires dialectales" (310; cf. Wagner 1959:108).

Schmidt sums up: "Es ist wohl auch beim heutigen Stande der Wissenschaft noch unmöglich, in diesen schwierigen Fragen [of macro-typological comparison] zu einer Evidenz zu gelangen" (351) --- a sentiment repeated a quarter century later in his necrology of Wagner (1989:228).

Jackson criticizes Wagner's sprachgeographical approach from a rather different angle: for taking a "synchronic approach to what is properly a historical problem" (1962:230),¹²⁶ and for neglecting issues of chronology. He condemns (233) Wagner's explicit ahistoricity --- significantly, not only as regards Wagner's treatment of the macro-geographical "Eurafrican" stratum, but also of the British verb itself. If native, pre-Indo-European syntactic patterns were truly influencing the evolution of the British languages, an influence which could emerge into the written record only when a previously dominant Hochsprache collapsed, then why is it that the collapse of Old Irish into Middle Irish does not involve a notable increase in "British" features? Do Middle English, or early modern English or Welsh, for which "there are plenty of written sources ... that are more or less vulgar and dialectal," exhibit a rich development of British features? "The answer ... is quite often 'No'," Jackson observes, "yet these questions are not even asked [by Wagner]" (231). Jackson also criticizes what he sees as Wagner's overreliance on modern forms of the languages, and castigates Wagner for not sufficiently considering the possibility that putatively "British" features in Celtic languages may reflect nothing more than the influence of English. This is especially so for Manx and Cornish;

¹²⁶ Similarly Gagnepain: "Il semble a priori dangereux de conclure de la typologie à l'histoire" (311).

whereas Wagner (102) presents these as primary manifestations of the "Eurafrican" stratum, Jackson refers to modern-day Manx as a "grossly broken-down jargon, half-forgotten ... and thoroughly penetrated with English influences," and similarly for late Cornish.¹²⁷ "To use these jargons as evidence on ancient insular Celtic and its relations with 'Berber' is a defect of method" (231). Finally, he criticizes Wagner's almost total silence regarding Breton: "the 'Berber' influence on the British from which Breton derives must have been much older than the separation of Breton, and Breton should therefore show most of the 'British' features of Welsh and Cornish" (232).

Jackson's concern for chronology informs his entire career (see especially Jackson 1953), and his remarks must be taken very seriously. But the points he raises here tend to be stated oversimply and overzealously. Regarding pre-Celtic substratal influence, first of all, Jackson's answer to his own rhetorical question (above) is not a categorical "no" but a qualified "quite often no", which also means "sometimes yes". It is hard to escape the implication that, for Jackson, a substratal explanation cannot be valid unless it answers to the same standards of exceptionless rigor that would be demanded of (say) a sound law, and applies more or less perfectly in every documented case. Such a standard of methodological purity, however, is utopian and inappropriate when dealing with contact phenomena between languages. The Semitic languages of Ethiopia, for example, have all been syntactically

¹²⁷ Wagner is aware of the problems here, but emphasizes that the English influence on Manx "kaum die grammatischen Hauptkategorien betrifft" (1959:89). Later he berates Jackson for "mangelhafter Feldforschung" ["deficient fieldwork"] on Manx (1964:293).

reshaped under substratal influence, yet that reshaping has occurred in different ways, to different degrees, and presumably at different periods in the various languages (Greenberg 1980).

Regarding English influence on the modern Celtic language type, Manx and Cornish in fact figure in Wagner's presentation first and foremost for intra-Celtic comparisons. The bulk of the concrete "Eurafrican" comparisons involve not these "degenerate" dialects but Welsh and especially Old Irish --- languages which are chronologically prior to even the earliest Modern English. Wagner's "defect of method", then, is less severe than Jackson would have it be. Of course Wagner also makes heavy appeal to modern dialects of Irish and Welsh, as Jackson points out (231), and these might both seem suspect of English influence. But these are hardly "broken-down jargons". As for Irish, it is Irish influence on English (Anglo-Irish) which seems patent rather than the reverse, especially for the verbal features Wagner is examining. For Welsh, Wagner argues that the syntactic type of modern spoken (not literary) Welsh is near-identical to that of Middle Welsh,¹²⁸ a language which predates modern English. The syntactic influence of English on modern Welsh and Irish, then, may not be nearly as serious a problem as Jackson implies --- especially if our focus is (as here) specifically on Celtic/Hamito-Semitic comparison.

¹²⁸ "Unsere Untersuchung stützt sich deshalb auf den gesprochenen Dialekt und auf die Sprache der mittelalterlichen Prosa, die einander typologisch so nahe stehen, dass man von einer Sprache reden darf" (1959:60). "Spoken Welsh, although riddled with loan-words from Modern English, [is] in various ways closer to and, therefore, a more genuine descendant of the mediaeval language than standard Modern Welsh" (1981:54) --- a view, Wagner adds, which was shared by David Greene.

Of all the issues dealt with in the book, it is perhaps Wagner's comparison of Insular Celtic and Hamito-Semitic that evokes least in the way of specific concrete response. This is not surprising, given that the critics (some by their own admission: Schmidt 350, Jackson 232) lack the requisite familiarity with both language groups. There is little dispute that "Wagner has compiled an impressive list of quite striking apparent [sic!] similarities between insular Celtic and (particularly) Berber" (thus Jackson 232; similarly Sommerfelt 229), or that he is too bold with his conclusions. But only Sommerfelt and Jackson venture beyond generalities; and almost the only specific point that comes up is the etymological connection Wagner draws between Irish and Berber -d- and -n-. Both critics are puzzled. "One does not know in what way such elements could be transmitted," says Sommerfelt (231). Jackson sees what Wagner is up to somewhat better: "It is perhaps intended that the early insular Celts made use of certain IE. elements in new ways under the influence of some neighbouring (rather than substrate?) language of Hamito-Semitic type" (233). But were Berber -d- and -n- (and reciprocal -m-) actually d, n, m millennia ago, Jackson asks? "This seems doubtful" (233); it is a question Wagner ought to raise but does not, and one which for a historical linguist is fundamental.

Several methodological points in conclusion. To no surprise, Wagner's terminological framework is attacked for vagueness (Schmid 1960:313-14, Schmidt 1963:349, Rössler 1960:142, and especially Gagnepain, who devotes half of his review article to terminology [1961:314ff.]).¹²⁹ His correlation of linguistic types with

¹²⁹ Schmidt: "Eigenwillig und unklar scheint mir z.T. die Terminologie von Wagner zu sein"; Rössler remarks on Wagner's "mehr verschleiende

anthropological types, "Geistesstruktur" (1959:241-42), or national Weltanschauung (44, 50, 145) is just as unpopular: Jackson (233), Gagnepain (314), Schmidt (348), Sommerfelt (231) --- such views are "very daring" in today's linguistic climate, as Sommerfelt diplomatically puts it. Two critics also point out an overreliance on translation equivalents. Thus Sommerfelt: "The author is inclined to classify Celtic elements according to how they are translated into the usual European languages" (231) --- for example, in speaking of the "verbal" character of the conjugated preposition in Irish phrases like

is maith liom "I like it" (lit. "is good with-me")

(cf. Wagner 1959:169). Similarly Gagnepain: "Il est à craindre que le départ soit, consciemment ou non, fonction de la traduction" (319).¹³⁰ Finally, and perhaps most significantly, Gagnepain dissects Wagner's repeated claims that this or that element is "verbal" in nature (see sec. 2.4.1): Wagner has confused true "verbality" with the distinct notion of the "phrastic" (= predicative [?]) element in a sentence.¹³¹ "Point n'est besoin par conséquent, et même tout interdit, de supposer je ne sais quelle 'verbalité' de l'adjectif (p. 191, 223) ou des préverbes (p. 167 sq.), ce qui est contradiction dans les termes" (324).

als erhellende Terminologie." Gagnepain's discussion is marred by technical jargon that is just as idiosyncratic and off-putting as Wagner's.

¹³⁰ Ironically, Wagner himself is also aware of the pitfalls involved in overreliance on translation (1959:183).

¹³¹ An element is "phrastique" if it functions "en tant que signe de la phrase" (324). I take this as meaning the element which carries the sentence's predication force.

2.5 Other recent work

2.5.1 G. B. Adams

In recent decades the only substratalist other than Pokorny and Wagner to maintain an ongoing research interest in the Celtic/Hamito-Semitic problem has been G. B. Adams (1956, 1970, 1975, 1980)¹³² --- with little, unfortunately, in the way of new concrete results. Adams's primary concern is not typology but place names. His 1980 article in particular lays out the methodological issues involved in trying to ferret out Mediterranean substratal elements in Celtic vocabulary --- especially in place names or in the many Old Irish words lacking a good Indo-European etymology. He contrasts the different phonological systems of Basque, Berber, Irish, and Welsh, discussing how these differences might have altered the shape of possible borrowings. Elsewhere (1956:18) he suggests that the alleged slowness of phonological change in Hamito-Semitic potentially augurs well for a hunt for possible substratal loan-vocabulary: "There should still be a chance of recognising the connection between loanwords in Gaelic from such a source and their cognates in known Hamito-Semitic languages." Phonological conservatism is indeed a hallmark of Arabic; however, elsewhere (contra Adams) drastic changes can occur. Thus Akkadian shows near-complete loss of laryngeal consonants, Hebrew develops a new stop/fricative alternation, Aramaic and Akkadian show heavy loss of unstressed short vowels, Ethiopic has collapsed short *ĩ* and *ũ* into schwa, etc.; as for Hamito-

¹³² The 1956 article appears to be unavailable in American libraries. The 1975 study is unique, among all articles ever written on the subject, in having been presented at a conference on Hamito-Semitic.

Semitic, so radical have the changes been that the establishment of inter-branch cognates is still largely a matter of hunch; the evolution of Egyptian into Coptic involves major transformations; and attested borrowings of Semitic and Egyptian words into Greek and Latin show how drastically such loanwords (especially as regards pharyngeals and emphatics) can be deformed.

•
Structural similarities between Celtic and Hamito-Semitic come up in passing, largely as reiterations of points made by Morris-Jones and Pokorny. Adams emphasizes particularly the co-presence in Celtic and Semitic of the "construct chain" of embedded genitives --- the type

key door house teacher the-school

"the key of the door of the house of the teacher of the school"
--- and of frequent lexicalized combinations of the sort "son of X"
("son of death" = "criminal") (1956:11-12, 1975:240-45).

Adams touches on other points. Most of the articles discuss the chronological issues involved in substratal transmission, including capsule indications of the archaeological evidence; also, the various possible source languages, corresponding to the various migrations to Britain, are catalogued (see especially 1980:49-53). There is also a brief comparison of similarities in the grammatical traditions of Irish and Arabic (1956:5-7, drawing on Bergin 1938b:209, 212), with a suggestion that medieval Irish grammarians might have known of Hebrew grammatical works; this is followed (8-9) by a good summary of the fantastical Celtic/Near Eastern links proposed in the 18th and 19th centuries. In a more theoretical vein, Adams discusses why a posited Hamito-Semitic substrate language, in its slow advance toward the British Isles, should

have left almost no continental traces: "The tides of Indo-European migration were stronger and more numerous" in western Europe than in the backwater of Ireland, with successive waves --- Continental Celtic, Latin, Germanic, and "educated" Latin --- each overlaying and blurring any possible Hamito-Semitic residue on the continent (1956:13). As for the long period of coexistence alleged for pre-Celtic and Celtic languages in Britain before the final ascendancy of Celtic, Adams points out that "the English language has existed in Ireland for some eight centuries but it is only within the last century and a half that it has become the majority language of Ireland and that Irish has suddenly shrunk away to the remotest corners of the country" (1980:53). Long-term Celtic/pre-Celtic bilingualism is thus not at all unreasonable.

2.5.2 Karl Horst Schmidt

The most recent contribution to the problem is Schmidt 1990, a contrastive substratal study of Tocharian and of Insular Celtic. The paper presents nothing new factually, but is interesting for its methodological points (and errors). Earlier in his career Schmidt's attitude had been one of careful skepticism, notably in his 1963 review of Wagner's Das Verbum: "Es ist wohl auch beim heutigen Stande der Wissenschaft noch unmöglich, in diesen schwierigen Fragen zu einiger Evidenz zu gelangen" ["It is probably still impossible, in the present state of {linguistic} science, to attain to any certainty in these difficult questions"] (1963:351, repeated 1989:228; see also 1976:54-55). In his 1990 article Schmidt distinguishes between (1) singling out features of Insular Celtic which are likely to be substratal, and (2) assigning a concrete

identity to that substratum. That there was substratal influence of some sort he takes as established: "[Insular Celtic] must have had contacts with pre-IE substrates. In Celtic, the differences between Continental Celtic ... and Insular Celtic ... may be partly explained as the result of a pre-IE influence on Insular Celtic" (1990:196). As regards the identity of the substratum, however, he continues to be skeptical: "The identification of non-IE substrate languages is ... quite impossible for Insular Celtic" (195, cf. 1989:228). Morris-Jones's basic postulate, that "neo-Celtic syntax agrees with Hamitic on almost every point where it differs from Aryan" (1900:618), has proved oversimplistic. Nonetheless, Schmidt's dictum "quite impossible" ignores the fact that, of the typological isoglosses linking Insular Celtic to other language groups, a strong majority point specifically to Hamito-Semitic (and not, say, to Basque); see sec. 2.4.2.3. Schmidt further objects that the alleged pre-IE features of Insular Celtic are all syntactic, and "it stands to reason that morphological evidence such as the appearance of agglutinative inflection in Tocharian carries more weight than syntactical and semantic structures" (189). I do not understand the reasoning. Morphological resemblances involving sound and meaning are of course of unique importance in genetic argumentation, but in purely structural, nongenetic comparison I see no reason why morphology should enjoy any preeminence over syntax.

Schmidt's primary concern, however, is the identification of those features of Insular Celtic that are likely to be due to a substratum --- any substratum. He presents a rather arbitrary selection culled from Pokorny, Wagner, et al., rejects some, and settles on three as non-IE in type: "Basic Word Order VSO, conjugated prepositions and an increasing

tendency towards periphrasis" (195). Regarding VSO word order, he makes the important point that Insular Celtic deviates here not only vis-a-vis Indo-European but vis-a-vis Continental Celtic (Gaulish SVO, Celtiberian SOV). In light of the fragmentary state of our knowledge of Continental Celtic, word order is practically the only Insular Celtic structural feature which we can compare to an earlier stage of Celtic, a fact which makes this feature particularly precious in substratal argumentation (191). A further argument in favor of substratal influence, Schmidt suggests, is "the opposite direction taken by the transformations in Tocharian and Insular Celtic" (190).

Schmidt attempts to rebut a number of the often-cited Celtic/Hamito-Semitic parallels, appealing to a variety of rather unsatisfying and traditional arguments. Several of the features, he suggests (apud Thurneysen 1930:428), might have been influenced by colloquial language, e.g. (187-88):

- a) Resumptive pronouns after prepositions in relative clauses;
- b) The omission of the copula;
- c) An "amplified" negative construction à la French "ne...pas";
- d) Juxtaposition of clauses without subordination.¹³³

Perhaps; but there is no a priori reason to think of these features as natural "colloquialisms", witness the fact that (a), (b), (d) are

¹³³ The list is Schmidt's; Thurneysen himself makes only the general point, suggesting (with no particulars) that some of the "special" Celtic features might well be found in other old Indo-European languages if we had access to their "Umgangssprache" as well as their "Literatursprache". Point (c), in any event, is well-known as a recurrent development in Indo-European, and hardly peculiar to Celtic; see Meillet 1965[1912]:140.

ancient features in Semitic and Egyptian, and quite independent of register (the Bible and the Koran are hardly "colloquial"). Schmidt also criticizes the view that initial mutations are something peculiarly Celtic and un-Indo-European; as a sandhi phenomenon, there is nothing special about mutation (189). But this is beside the point. The issue is not the sandhi origin of the mutations but their functional recasting within Insular Celtic, where they have cut loose completely from phonology and become a system of syntactic marking via word-initial change --- "ein absolutes Unikum" within Indo-European (Wagner 1959:121).

Finally, Schmidt makes an extremely important observation in his 1989 necrology of Wagner, noting that "das Anwachsen dieser 'unindogermanischen' strukturellen Züge im Verlaufe der Sprachentwicklung einer zusätzlichen Erklärung bedarf" ["the growth of these 'un-IE' structural features over the course of the language's development calls for additional explanation" (italics mine)] (1989:228). Even if, for Pokorny and Wagner, such growth was part and parcel of the substratal or sprachgeographical explanation, it is not a necessary part of a substratal argument. It is easy, and natural, to envisage substratal influence operating only at the early stages of Insular Celtic prehistory, prior to the emergence of Old Irish and Welsh. Such influence could have set Insular Celtic in the same general mold as Hamito-Semitic; subsequent developments in Celtic (and Hamito-Semitic) might then reflect the natural drift of languages of this type.

2.5.3 Other studies

Other scholars have made individual contributions to the problem of Celtic/Hamito-Semitic resemblances. D. S. Evans 1950 provides a concise and clear statement of syntactic parallels between Hebrew and Welsh. Hewitt 1984-85 does the same, in great detail, for Breton and Arabic, drawing in part on Greenbergian word order typology. Neither author goes beyond the two languages in question, or addresses the question of the origin of the similarities. Salvaneschi (1978) proposes to account for certain Celtic-Semitic resemblances by appealing to biblical Latin as textual intermediary. Hans Hartmann (1954, 1960) examines several characteristically Celtic constructions from a strongly Humboldtian (even Finckian) viewpoint, sometimes verging on the mystical,¹³⁴ but with no reference to Hamito-Semitic. Harald Haarmann, in his book-length study of areal typology and European Sprachbunds (1976), summarizes Wagner's views on the British type/Sprachbund (117-22) and on the Eurafrikan type/Sprachbund (135-39); for "Eurafrikan", he notes, the term "Sprachbund" is not really appropriate, as the speech communities in question are not in contact --- the observed typological convergences are rather to be ascribed to independent parallel development (138).

As far as I know there has been only one attempt thus far to treat the problem from a purely typological, nonhistorical and nonareal perspective. In a programmatic article (ms), the Semitist K. Jongeling lays out a detailed list of structural similarities between Hebrew and

¹³⁴ Thus Hartmann proposes to elucidate Irish impersonal constructions by invoking a quasi-magical "Allkraft" (e.g. 1954:34), a turn which Dillon (1955:109) characterizes as a "lapse ... into mere fantasy". See further 1954:104 for Hartmann's Finckianism ("Es handelt sich in Irland offenbar um eine spezifische Denkform"), and 1954:15 for a romantic appeal to the fairies in explanation of the impersonal.

Welsh (12-18), summarizes briefly the previous substratal and areal tradition of analysis (1-5), and turns to the possibility of a typological approach. "When it can be shown," he says, "that typological characteristics found both in Hebrew and in Welsh are common to many languages all over the world there is no need to find a special link between the two languages presently under discussion" (7). If the shared features are recognized as those characterizing an established linguistic type, then the question of any historical link among the languages becomes moot: typological relatedness itself suffices as explanation. This would be very satisfying as an explanation, for then the absence of similarities in vocabulary or phonology would become a non-issue. But, as Jongeling notes, we do not know whether the cluster of Celtic/Hamito-Semitic features does in fact form a type. Word order typology recognizes an established clustering of "ideal VO" features, but the similarities at issue go far beyond word order.¹³⁵ A detailed investigation into the typology of VSO languages would be a signal contribution toward solving this question (9). This would, however, leave unsolved (even unaddressed) the question of "why insular Celtic chose a way of development [VSO] that is not in line with the general trend in the Indo-European languages" --- a developmental path which is, moreover, at odds with the crosslinguistically favored SVO and SOV (10).¹³⁶

¹³⁵ Other scholars have argued, both empirically and theoretically, that there is a natural typological link between verb-first word order and non-word-order properties such as the presence of verbal nouns (thus Jeffers 1976, Myhill 1984 [Chapter 6 on verb-initial languages, esp. pp. 221ff.]). These studies, however, make no reference to the CHS problem.

¹³⁶ See Jongeling 1991 for brief discussion of VSO vs. SVO orders in Welsh and Hebrew.

Chapter 3: The roster of features surveyed

3.1 Methodological requirements

A great many points of syntactic resemblance between Celtic and Hamito-Semitic were proposed in the previous chapter. Obviously, not all of these "CHS features" have equal validity; many in fact verge on substratophile wishful thinking. In deciding which features to examine in the typological survey to be presented in the coming chapters, I have therefore followed a number of conservative but not rigid guidelines:

1) Language distribution:

a) The feature must occur in either Old Irish¹ or Middle Welsh. I do not insist on Old Welsh because the corpus is extremely small. I choose Welsh over Breton largely because the Welsh medieval corpus is so much larger, and the language so much less exposed to superstratal influence (French). Further, insular substratal influence on Breton could only have manifested itself at a period prior to the emigration of pre-Breton to the Continent; any such feature should thus also exist in Welsh (unless subsequently lost). (Conversely, a Welsh feature ideally should also exist in Breton.) Of course, Old Irish is in general more archaic

¹ Some of the "Old Irish" examples to be cited actually come from the saga literature, which we have only in Middle Irish redactions. These examples are taken to be representative of "real" Old Irish (as found in the glosses).

than any attested Welsh;² but it seems unreasonable simply to exclude out of hand one of the two extant sources of early Insular Celtic, which are far from structurally identical. And substratal influence need not have been the same on the two islands; witness the situation in Ethiopia, mentioned in sec. 2.4.4, where Cushitic substratal influence manifests itself to different degrees in the different Ethiopic Semitic languages.

I am, however, specifically excluding features which appear only in the modern Celtic languages. Any such parallels involving the modern languages I will provisionally dismiss here as the result of drift (but see sec. 7.4 for further discussion of such chronological anomalies). The main reason for this focus on the older languages is straightforward: despite Wagner's views about the transtemporal nature of linguistic areas, I cannot see any possible path of transmission for substratal influence by the time Modern Welsh (~1400) and Modern Irish (~1200) came into existence. The substratal language would have had to still be spoken, and this seems unlikely.

It may not be out of place here to lay out the argument in more detail than was given in sec 2.3.1.2. It is, of course, not known by what date the substratal language died out in Britain and Ireland. For Welsh, this can perhaps be inferred indirectly from the history of Pictish, the only substratal language and people mentioned in the historical record.³ The Picts, whose kingdom was in northeastern Scotland, lost

² Though note Greene's comment: "I cannot think of any feature of Irish which is to be considered an archaism and of which no trace is to be found in British" (1983:136-37).

³ The Venerable Bede, who died in 735, mentions four languages and na-

their political independence in the ninth century (Wainwright 1955:4); by the mid-twelfth century Pictish was in all likelihood a dead language,⁴ at least in southwestern Scotland (1955:40-41) and a fortiori further south. If so, it could no longer have been in contact with Welsh. The mid-twelfth century marks the beginning of the Middle Welsh period (Evans 1976:xvi-xvii); thus any substratal influence must have ended long before Modern Welsh began. Of course, there is no guarantee that the entity we call "Pictish" was the relevant substratal language. But historical sources, as far as I know, mention no other pre-Celtic group extant in medieval Britain; it would seem that the Picts were the only "alien" group to have registered on the consciousness of British writers of the time. Had some other such group survived up to the twelfth century the silence in the historical record would be strange.

--- An analogous argument might be made for Ireland, but with much less confidence, for here we have even less information than for Britain on the pre-Celtic peoples of the island. Again, the argument depends on silence. As G. B. Adams puts it, "By the time Irish literature begins in the 6th/7th century A.D. ... there would appear to be no surviving tradition of any pre-Celtic language being spoken in Ireland" (1980:52). And even if a marginalized pre-Celtic language were still spoken in isolated pockets of the country, its status would have waned over the coming centuries. The beginning of Modern Irish is set roughly at 1200 (Greene 1966b:12), over half a millennium later --- a very long time for an "invisible" substrate language to survive without a trace.

tions in Britain: Britons, Picts, Scots, and Angles (Wainwright 1955:3).

⁴ "According to Henry of Huntingdon, the Picts and their language were so completely destroyed [by the twelfth century] that a mention of them in earlier writings [Bede] seemed like a fable" (Wainwright 1955:40).

Leaving modern Celtic out of consideration has a serendipitous consequence. It is a happy accident that a salient Insular Celtic trait pointing not to Hamito-Semitic but to Basque --- the tendency to "Flexionsisolierung" --- is explicitly a late Celtic feature. It would be important to look in detail at Flexionsisolierung if Basque did in fact share many old Celtic/Hamito-Semitic traits; but (as we will see in Chapter 6) it does not. Limiting the investigation to early Insular Celtic thus naturally --- and appropriately --- helps keep Basque at arm's length.⁵

b) The feature must occur in one of the branches of Mediterranean Hamito-Semitic: Modern Berber, Ancient Egyptian, ancient Semitic. For Egyptian I focus almost exclusively on Middle Egyptian; it would be unwise to base inferences on Coptic, which is attested so late (3rd century AD [Gardiner 1957:5]) and has undergone such radical re-formation (see sec. 6.5). Modern Berber, of course, is not an ancient language at all. I use it only because we have no earlier form of the language; the Numidian inscriptions representing ancient Libyan (roughly the time of the Romans), though probably ancestral to Berber, constitute a tiny corpus and are not well understood (on Numidian see e.g. Rössler 1958).⁶ If a feature is cited only for Semitic, ideally it should occur

⁵ Recall Mac White's plea to clearly delineate and separate Hamito-Semitic vs. Basque substratal influences (1955:12).

⁶ Berber is not a single language but a large cluster of closely related dialects. I will focus on the Kabyle dialect (Mediterranean coast of Algeria), as described in Chaker 1983, with occasional reference to other studies (and other dialects) when necessary. In so doing, I am assuming (and I think the assumption is a fair one) that the Berber dialects are relatively homogeneous in overall syntactic structure; witness Basset's introductory remark to the chapter "Morphologie et syntaxe" in *La langue berbère*: "Les principes généraux sont valables pour tous les parlers" (1952:11).

throughout the family, or at least in several languages. I exclude Chadic and Cushitic from the CHS languages partly on geographical grounds, but primarily because the full-fledged CHS type appears not to show up in these languages. Both groups have themselves been subject to heavy areal and substratal influence, distancing them from the Mediterranean type.

Note that the only modern language appealed to here (whether Celtic or Hamito-Semitic) is Berber. Some attention ought properly to be paid, therefore, to the likely antiquity of the various Berber features. In practice, this question becomes ensnarled in deep and unresolved issues of syntactic reconstruction, and answers are seldom forthcoming. I will in fact raise the point only once, with regard to word-initial change (feature [16] and sec. 7.4).

c) The feature must not be a dominant feature of old Indo-European languages. It can occur in some languages, ideally only in scattered ones, but not in most. The feature can show signs of being an archaism within Indo-European (notably as regards Celtic-Hittite points of agreement); as argued, substratal pressure can seize on an inherited feature and foster its development at the expense of other structural alternatives. What is important is that the feature should be clearly "alien" to the bulk of the family, for otherwise there would be no reason to look to a substratum in the first place.

d) Ideally the feature should not occur in Continental Celtic; even more ideally, its opposite should occur --- e.g. V-initial order in Insular Celtic vs. non-initial order in Continental Celtic (SVO or SOV, both better representing the old IE type). However, the same caveat as

for Numidian (above) holds here: we know too little of Continental Celtic, and some of the major inscriptions are still baffling in many respects. Hence one cannot insist on a disagreement between Continental and Insular Celtic. On the basis of currently known data, none of the CHS features occurs in Celtiberian, which is considered the most conservative form of Continental Celtic; several of the features do occur in Gaulish, on which see discussion in sec. 7.4.

2) In the Celtic and Hamito-Semitic languages selected, the feature must be a clear and dominant characteristic of the language, not a trace phenomenon or a "minor" pattern.

3) The Celtic feature and the Hamito-Semitic feature must resemble one another very closely as regards both syntactic patterning and meaning and function. This criterion would exclude, for example, any Berber-Basque-Irish comparison of elements such as d or n, which show only a general resemblance of meaning and function.

4) The feature must be an obvious "surface" feature of the languages. There are two quite different reasons for this. First, "deep" features are always suspect of being artifacts of the deep analyst's theoretical orientation (as was so often the case for Wagner). The skeptic may doubt their reality, and sometimes with good reason. Secondly, there is the purely practical problem of identifying the presence/absence of the feature using conventional reference tools. Asserting, for example, with Wagner that Insular Celtic and Berber preverbs are "verbal" in nature (quite apart from whether the claim is in fact correct) requires a profound "feel" for linguistic structure which will seldom be forthcoming from reference books.

5) The feature must be statable in a precise way. Cataloguing languages' overall "nominal" or "verbal" character --- even if it were doable --- would be too vague to be useful so long as these notions remain inexact.

6) The ensemble of features need not convey any impression of typological coherency; they can perfectly well be a hodgepodge, and even trespass into the lexicon (as with the "KIN of" construction). The point at issue is precisely that we do not know what features "ought" properly to cluster together.

7) Every feature must be potentially specifiable as an opposition set. That is, if Celtic and Hamito-Semitic agree on a feature value [+X] or [alpha-X], it must be clearly statable what other values the variable X may assume along the given typological dimension. I defer treatment of this very difficult question to Chapter 5, where it will be dealt with in a much broader perspective.

The present chapter is devoted to a statement of the relevant facts as found in Hamito-Semitic, Celtic, and Indo-European. Accordingly, each of the features catalogued below will be presented according to a standard format:

- a) As found in Hamito-Semitic.
- b) As found in Insular Celtic.
- c) As (not) found elsewhere in Indo-European.
- d) Comments, if any.

Often the "Comment" will aim at countering this or that argument to the

effect that a given Celtic feature "could not", on theoretical grounds, be ascribable to substratal influence. At issue in such cases is only the logic of the argument --- not whether the feature really did arise under substratal influence, only whether it logically could have done so.

It should also be emphasized, especially for Indo-European, that the features are considered here strictly in terms of their synchronic functioning within the given languages, not with regard to diachronic origin. Finally, a note on sources: many of the examples given below, especially those illustrating obvious and uncontroversial features of Welsh and Semitic, are presented schematically using invented data. Other examples will include a brief reference to source.

3.2 Features: Hamito-Semitic, Celtic, Indo-European

1) Conjugated prepositions (Prep-suffix)

a) Found throughout Hamito-Semitic (Arabic min-ka "from-you"). The syntagm is structurally identical to Noun + pronominal possessor (Arabic baytu-ka "your-house").

b) Found throughout Insular Celtic (Old Irish fri-t "against-you"). A few prepositions in Welsh instead take independent pronouns as object (ac ef "with him"), but they are a small minority (see the list in Williams 1980:129-30 for Modern Welsh). Old Irish and Welsh also have denominal prepositions which are still transparently nouns, and as such

take their object as proclitic possessive pronouns (Old Irish far ndochum "your-toward", i.e. "toward you"); these again are a minority.

--- NB: In contrast to Hamito-Semitic, there is no parallel with the syntax of noun possession: possessed nouns do not take their pronominal possessor as a suffix, but as a proclitic.

c) Unknown elsewhere in Indo-European except in Hittite, where conjugated postpositions occur as an option, especially in older Hittite (Schmidt 1982:358). The postposition in the Hittite syntagm [Postp+PronObj] is construed like a neuter noun in a frozen case form, either Dat/Loc or Nom/Acc; this "noun" is followed by an ordinary possessive enclitic, which (as always in the older language) agrees in case with its Head (Friedrich 1974:65-66, 133-34):

	piran-tet	"before you" (Adp and Poss in Nom/Acc)
vs.	katti-ti	"near you" (Adp and Poss in Dat/Loc, ending -i;
		cf. <u>katta</u> "near", a different case)

Better paraphrases might be "[your-beforeness] (Nom/Acc), [your-nearness] (Dat/Loc)". As Wagner points out (1972a:69-74, 1985:66ff.), these constructions are closer in structure and spirit to typical OV syntagms of (e.g.) the Finnish type:

inside-at-its	"in it", (i.e. "at its inside")
---------------	---------------------------------

Nonetheless, the existence in Hittite of the syntagm [Adp + Clitic] makes it possible that something of the sort could already have been a syntactic option in PIE; if so, it could have been seized upon by Insular Celtic under substratal pressure, and thus become the norm.

d) Comment: In Middle Welsh the personal endings of conjugated prepositions are very similar to subject endings on the verb. By contrast, in Semitic and Berber the prepositional endings do not particularly resemble verb subject endings, but are identical or near-identical to noun possessive suffixes. (In Egyptian a single paradigm covers all three functions.) Such a morphological discrepancy might be thought to militate against any possibility of substratal influence in the development of Celtic conjugated prepositions. --- But the argument is weak, on two counts. First, the verb-preposition resemblance is not pan-Celtic but specifically Brythonic (the Irish prepositional endings show no close link to verb endings), and relatively late even there; comparing to Breton and Old Welsh, we can see the Welsh prepositional paradigm reshaping itself analogically after the verb (cf. Evans 1976:58). Secondly, Celtic, like the rest of old IE (except Hittite), does not have noun possessive endings at all. Etymologically, the paradigm of prepositional endings doubtless originated in IE enclitic pronouns;⁷ if this new paradigm were then to undergo any analogical re-formation at all, it could only have modeled itself formally after the paradigm of verb endings, the single analogical source available to it. The role of a hypothetical substratum would have been to foster the syntagmatic fusion of preposition and enclitic pronoun in Celtic, not to make the Celtic paradigm look like the Hamito-Semitic in points of detail.

⁷ My thanks to John Koch and to Stefan Schumacher for clarifying to me the morphological evolution of the prepositional endings.

2) Ideal VO macrotype: dominant VSO, N-Adj, N-Gen, N-RC1, Prep⁸

a) At the NP-level, ideal VO typology holds rigidly throughout Mediterranean Hamito-Semitic:

Akkadian: $\bar{b}itum\ dannum$	"house mighty" (= "mighty house")
$\bar{b}it\ ilim$	"house (of) a/the god"
$\bar{b}itum\ \check{s}a\ [\bar{i}pu\check{s}u]$	"house that [he.built]"
$ana\ \bar{b}itim$	"to a/the house"

The only exception is Ethiopic Semitic, where some OV patterns at the NP-level are beginning to emerge to various degrees even in Classical Ethiopic (Geez), under Cushitic substratal pressure (Gensler 1989): in Geez postpositions are unknown, Gen-N and RC1-N are minority patterns, while Adj-N is not uncommon. In modern Ethiopic, OV typology has become the norm.

⁸ The strong word-order correlation between VSO, N-Modifier, and prepositionality has been a commonplace in modern linguistics since Greenberg's study of word order universals (1966a), though the idea was already familiar to 19th-century typologists. I refer to this constellation as the "ideal VO macrotype"; its counterpart, the ideal OV macrotype, reverses the polarity of the above orderings (SOV, Modifier-N, postpositionality). None of the vigorous debate over the degree to which the macrotype should or should not be taken as a universal (see e.g. Comrie 1989, Chapter 4) affects its validity as a clear and strong tendency. --- In discussions of word order, I will be using the term "NP-level" as roughly synonymous with "phrase-level", intending thereby to set it off as the opposite of "clause-level". Including Prep phrases under this header may technically amount to a mixing of levels, since in a PrepPhr not the N but the Prep is the phrasal head. In many language groups, however, and notably Semitic, the syntagm [Prep Obj] is transparently the same as [HeadN DeptGen]; and the word-order correlation between the two is probably the strongest of all pairwise correlations of word-order parameters. Hence it does not seem out of place to lump PrepPhrases in with N-modifier constructions as constituting a single level of word-order parameters.

At the clause level, VSO order is standard but not rigid in Egyptian, Berber, and most old Semitic languages:

Hebrew: va -yar elohīm et ha-ʔor "and God saw the light"
 and-saw God ACC the-light (Genesis 1:4)

Akkadian has SOV clause-level word order regularly, and Aramaic (under substratal pressure) frequently; thus:

Akkadian: Šumma awīlum awīlam ... imtaḥaṣ
 if man(nom.) (acc.) he.has.struck
 "if a man strikes a man" (Law Code of Hammurabi, 206)

But these languages are otherwise "ideal VO" in type (see above for Akkadian). Some word order flexibility at the clause level is common throughout Semitic for expressive purposes.

b) At the NP-level, VO typology holds fairly rigidly throughout Insular Celtic:

Welsh:	tŷ mawr	"house big" (= "big house")
	tŷ gŵr	"house (of) a man"
	tŷ a [wnaeth]	"a house that [he made]"
	i dŷ	"to a house" (with lenition)

Rare instances of lexeme-specific Adj-N order exist, e.g. Welsh hen "old"; otherwise Adj-N and Gen-N order, when attested, are analyzed as compounding (witness mutation).

At the clause level, VSO is standard in Old Welsh and Modern Welsh, and in Irish at all periods:

Old Irish: benaid Cú Chulainn omnae "Cu Chulainn strikes a tree"

(example RIA s.v. "omna"). Middle Welsh makes heavy use of SVO order, which in Breton has become the norm. Old Irish also has a constellation of rare OV phenomena (tmesis, Bergin's Law, also Gen-N) --- undoubtedly IE survivals --- which are characteristic of the archaic poetic language and (as argued by Wagner 1967, 1977) play a central role in poetic alliteration; see e.g. Bergin 1938a, Greene 1977.

c) At the NP-level, most older Indo-European languages show flexibility, allowing the orders N-Adj/Adj-N and N-Gen/Gen-N; adpositions are usually prepositions, but in many languages may also occur as postpositions or even as "interpositions", as in:

Latin: summa cum laude "with highest honor(s)"

 highest with honor

Relative clauses are predominantly of the postposed "headed" type (order N-RC1). Hittite, however, shows consistent OV typology: Adj-N, Gen-N, Postpositions; Hittite relative clauses are of the correlative type (see [4c] below). --- At the clause level, a similar pattern obtains in most of the old languages: flexibility prevalent, OV typically dominant, VSO distinctly in the minority; Hittite again is solidly OV in type (SOV, though with some flexibility). Specific mention should be made of Continental Celtic: SVO for Gaulish; SOV (and overall OV type) for Celtic, according to Schmidt (1976:55, 1990:190-91). --- To varying degrees, OV features recur throughout early Indo-European, notably in the oldest strata of the languages. PIE itself shows very high likelihood of having been of OV macrotype (Holland 1980); some scholars prefer to reconstruct a flexible word order (on the model of Greek); on either view, a VO macrotype is out of the question. VSO order does exist as a

marked order throughout old Indo-European, but only in Insular Celtic is it the norm.

3) Relative clause linker: no relative pronouns; relative particle(s) or zero

a) Relative pronouns ("WHich") --- that is, relative markers whose inflection shows the relative noun's function within the embedded clause --- are unknown in Hamito-Semitic. Rather, the languages make use of one or several clause-initial relative particles "REL" ("that"), whether invariant (e.g. Hebrew ašer), zero (asyndetic; direct juxtaposition), or inflecting to reflect gender/number and the noun's role in the matrix clause (Arabic allaḫi; case coded only in the dual).⁹ Thus:

Hebrew: ha-ʔiš ašer [raʔiti] "the man that I saw"

Berber: argaz (da) [tzrit] "the man (that) you saw"

(Berber example from Sadiqi 1986:33-34; the relative marker is optional.)

b) Relative pronouns are unknown in Insular Celtic. Breton uses a clause-initial relative particle a. Welsh has two relative particles, a for subject and object relatives, y for oblique relatives:

⁹ Note also that the linker-type in Arabic depends on the definiteness of the HeadN: zero with indefinite HeadN, allaḫi with definite.

Welsh:	y	gwr	a	[welais]	"the man <u>that</u> I saw"
	y	gwr	y	[gwelais ei dy]	"the man whose house I
	the man	REL	[I.saw	his house]	saw"

Old Welsh sometimes dispenses with a relative particle entirely. Old Irish RCLs show great complexity, but relative pronouns are never involved. The most typical pattern, found with prefixed verbs, involves mutation of the verb: lenition (subject relatives) or nasalization (oblique relatives) or either (object relatives). These mutations are reflexes of earlier relative morphemes, still visible as such in Continental Celtic --- independent relative pronouns in Celtiberian, invariant relative suffix in Gaulish (examples, [5b] below), both based on the IE element *yo- --- but they are not synchronically segmentable as a separate "linker" morpheme in Irish. A distinct relative particle per se can be found only for prepositional relatives, where it is invariant and "has the same form [- (s) aⁿ] as the acc. sg. neut. of the article" (Thurneysen 1946:312):

Old Irish:	di-a ⁿ	-dílgid	"to whom you forgive"
	to-REL-	you.forgive	

(Thurneysen 1946:28; see [4] below).

c) Elsewhere in older IE (including Celtiberian) inflecting relative pronouns (Latin qui/quae/quod), showing the role of the noun in the RCL, are all but universal as linker; only Old Norse (and no other old Germanic language) has an invariant relative particle, er (later also sem). The relative pronoun serves double duty, functioning both as linker and as (fronted) pronominal copy (see [4]). In Hittite, with its

correlative relativization strategy, the cognate element kuiš is not a "linker", but remains in situ in the RCl and acts as a relative adjective.

4) Relativization strategies: copying (resumptive pronoun); participles; special fronting strategy for prepositional relatives in Old Irish and Berber

a) Relativization by copying ("the house that you went to it") is standard in Semitic and Egyptian: a resumptive pronoun is required for oblique relatives, typically optional for object relatives, forbidden for subject relatives. Thus in Hebrew:

ha-bayit ašer [halaxta ēl-av] "the house that [you.went to it]"

Berber, remarkably, apparently never (or almost never?) uses resumptive pronouns. Semitic and Egyptian may also relativize using participles, both active and passive (or, in Egyptian, using special "relative" verb forms built on passive participles).

Berber relative clauses are of three types, depending on the function of the relativized noun; in all, the REL particle is optional. (Sadiqi 1986, Chapter 1, presents the phenomena with great clarity; examples below from pp. 38, 34, 40.) Subject relatives put the verb in a special "participial" form (not a participle in the usual sense), uninflected for person and featuring suffixed -n:

argaz (da) [isRa-n lktab] "the man (that) [bought the book]"

man REL buy-n book

Object relatives take the normal finite verb form, with no resumptive

pronoun:

argaz (da) [tzrit ____] "the man (that) [you saw ____]"

Prepositional relatives are formed by fronting the bare preposition, with (NB) no resumptive pronoun:

tamdint (da) [g izdR hmad ____] "the town [in which Ahmed lives]"
town REL in lives Ahmed

This last type is unheard-of in Semitic and Egyptian. There is an isolated example with fronted preposition in Phoenician (Segert 1976:171),

mkmm b-?š [kn ?šm r^Cm] "places [in which there were
places in-REL [were men bad] bad men]" (Karatepe I.14-15),

and similarly in Hebrew (Gen 31:32; see Gesenius 1910:446, who terms the construction "quite anomalous"); but even these do not match the Berber exactly.

b) In Welsh (similarly Breton), the basic relativization strategy is copying. A resumptive pronoun is standard for oblique RCLs (largely genitival and prepositional), nonexistent for subject RCLs; for object relatives, a resumptive pronoun is very rare with positive verbs, fairly common with negative (Morris-Jones 1931:92, 112). Example:

y gŵr y [gwelais ei dŷ] "the man whose house I saw"
the man REL I.saw his house

(Copying is also the standard strategy in Modern Irish, but this is beyond our purview here.) The Old Irish relative clause, with all its complexity (see [3b] above), does not prominently feature resumptive pronouns. They do not (apparently) occur with subject or object rela-

tives; they are found with certain types of genitival RCLs, e.g.¹⁰

fir as a c[h]athach "(of) the man that it is his trespass"
 man is.REL his trespass ("of the man whose trespass it is")

(examples, Thurneysen 1946:321-22), and occasionally in prepositional relative clauses (322). The standard construction for prepositional relatives, rather, involves no resumptive pronoun but fronts the bare preposition (1946:312); here (and only here) the invariant relative particle comes into play, and is positioned after the Prep:¹¹

di-aⁿ -dílǵid "(one) to whom you forgive"
 to-REL-you.forgive (Thurneysen 1946:28)

Nothing like this exists in Welsh. --- Insular Celtic lacks IE active (-nt-) participles entirely, except as opaque lexical vestiges; IE passive participles (-to-) survive as a full-fledged category only in Breton (and are usable there as relativizing devices), though reflexes exist in other functions in Old Irish.

c) No other IE language relativizes by copying, nor by fronting a bare preposition. The standard relative technique involves fronting an inflected relative pronoun (a "WH-pronoun"), leaving a gap in situ in

¹⁰ That the HeadNoun "man" in this example happens itself to be in the Genitive is an accident, with no significance for the construction.

¹¹ McCone (1985:96-97) proposes that the occurrence in Old Irish of two distinct strategies for prepositional relative clauses points to an ancient dialect split, the "Prep fronting" type being characteristic of the north and the copying type of the south, with canonical "Old Irish" accordingly based on northern dialects. The same geographical split exists today: whereas Modern Irish uses copying, Scots Gaelic preserves the old "Prep fronting" strategy. See also Ahlqvist 1988:28.

the clause;¹² thus in Latin:

domus [cui [uēnī ____]] "the house [to.which I came ____]".
house WH.dat I.came

Participial relativization is also very common. --- Hittite makes exclusive use of correlative relatives, a biclausal, nonsubordinating strategy involving relative adjectives (not pronouns): a noun is WH-marked as topic in the first clause ("WHich Noun"), then referred to anaphorically ("THat") in the second clause. To varying degrees this strategy is also found as a marked, minority option in most older IE. A Hittite example (Friedrich 1974:168):

nu-za d^uUTU^{ŠI} [kuin NAM.RA] INA É LUGAL uwatenun,
CONN Sungod WH.acc prisoner-group in palace I.brought
naš 15500 NAM.RA ešta
CONN-it was
"and [WHich prisoner-group] I, the Sungod, brought into the
palace, THat was 15500 prisoners"

To varying degrees this type, or its mirror image (with clauses reversed), is also represented in the other older IE languages (including Old Irish).

d) Comment: Here Insular Celtic and Hamito-Semitic show commonalities

¹² I will refer to this type of relativization as "gapping", the essential point being that a gap occurs in the clause at the expected site of the noun. ("Gapping" also has a very different use as a label for the phenomenon seen in sentences like "John likes football, and Mary ____ baseball", but this is not at issue here.)

with regard to two quite distinct relativization strategies, namely the copying strategy and the "fronted bare Preposition" strategy, both of which in turn are very different from anything found elsewhere in Indo-European. The agreement between Irish and Berber regarding the "fronted bare Prep" type is not perfect: Berber has the sequence REL+Prep, whereas Irish has Prep+REL; further, the REL element in Berber is optional, in Irish obligatory. But this does not vitiate the potency of the dual parallelism. It would be noteworthy for the two language groups to deviate from IE by employing a single shared strategy; it is all the more remarkable for this to occur twice over.

5) Special "relative form" of the finite verb (distinct from participles)

a) In Akkadian (Semitic), finite verbs in any subordinate clause take a special "subjunctive" form, with suffixed -u; this of course includes relative clauses:

bitum ša [īpuš-u] "house that [he.built]"

Such a phenomenon is unknown elsewhere in Semitic. Egyptian has a series of special "relative forms" of the finite verb (inflecting for person/number), usable only for non-subject relatives; these are morphologically linked to passive participles. Example:

bw wršw ib-i im "the place where my heart dwells"
place dwell.REL heart-my there (Gardiner 1957:300)

Berber has a special "relative form" (with suffixed -n) commonly termed a "participle", used only for subject relatives (Sadiqi 1986:38):

argaz (da) [isRa-n lktab] "the man (who) [bought the book]";

it does not inflect for person, and in some dialects is totally invariant. The term "participle" is a misnomer because the form has only a relative-clause use. (See Basset 1949 on the "participle".)

b) Simplex verbs in Old Irish have special relative endings in 3rd-person forms and in the 1-pl; contrast benaid "he strikes" with relative benas (example RIA s.v. "benaid"):

fer [benas dam]

"a man [who strikes an ox]"

(With other persons there is no special relative form; instead, a dummy preverb *no-* appears, converting the verb into a non-simplex form.)

These relative forms are apparently usable regardless of the syntactic role of the relativized noun in the RC1 (Thurneysen 1946:315, 319), with a single exception: the prepositional relative, which can never involve a special relative form because the fronted unit "Prep+REL" acts as a conjunct preverb and thereby renders the verb non-simplex. --- In Brythonic the single salient example of a relative verb form is the Welsh form ysydd "who/which is" (Evans 1976:63); the same relative ending -ydd is found in traces elsewhere in Middle and Old Welsh (1976:119). Continental Celtic (Gaulish) is problematic here, depending on the analysis of forms such as the famous dugiionti-io "they who serve(?)", toncsiiont[i]-io "they who swear", with their suffixed relative -io (cognate with Welsh -ydd and with one of the Old Irish relative endings). If such forms are taken as already fully unverbated (as indeed happened later in Welsh and Irish) they would count as a special relative form of the verb.

c) Relative verb forms are unattested elsewhere in old IE.

6) Polypersonal verb,¹³ coding object as well as subject

a) Semitic and Berber verbs can mark pronominal object as well as subject; Berber can standardly mark two copresent pronoun objects (both direct and indirect), a feature also to be found in Semitic (Arabic, Akkadian, Ethiopic), largely as an archaism. Thus:

Arabic: raʔaytu-ka "I.saw-you"

Berber: fki- γ -as - t "I.gave-him-it" (= "I gave it to him")
 gave I to.him it (Chaker 1983:138-40, 150)

Semitic object markers are exclusively suffixing; Berber object markers are suffixed to simplex verbs, infixes to complex verbs (see [7] below). Egyptian codes only subject directly on the verb; direct objects, expressed as clitics (of a special series), are separable from the verb by other clitics (Gardiner 1957:55, 186).

b) Old Irish verbs standardly mark pronominal object as well as subject; Object markers are suffixed to simplex verbs, infixes to complex verbs (see example under [7]). --- Welsh, too, routinely marks the object on verbs having preverbs, using infixes pronouns as in Irish; indeed, this is the common way of marking objects on non-imperative verbs (Evans 1976:50), given that all verbs normally take a grammatical preverb in most contexts. Only in a few early forms of the verb "to be"

¹³ The term is Wagner's (1959:155). Of course, what is coded on the verb is not just person but person/number of the relevant actants.

is there true suffixal marking of pronominal objects, in dative meaning: yss-ym "is-(to)-me" (142).¹⁴ With imperatives, pronominal objects are invariably marked with independent pronouns (49-50), and occasionally elsewhere; but this pronoun is freely separable from the verb complex, and hence is not part of Welsh verbal inflection. (This separability also holds for Modern Irish object pronouns.) In both languages, the pronominal object is usually the direct object, sometimes indirect; however, only one object affix is possible. --- Continental Celtic, it should be noted, already sometimes shows object clitics as part of the verb complex: Gaulish sioxt-i "he added them" (suffixal Obj), to-med-eclai "she set me up" (infixal Obj), and an additional example (Eska 1993:12).

c) This phenomenon is fundamentally alien to old Indo-European, where verbs inflect only for subject. There are several real and apparent exceptions involving clitic marking of objects. In Hittite, object markers are second-position clitics, and indeed can co-occur with subject clitics in the clause-initial chain of particles:

Particle.chain ... V #.

But only fortuitously, in this verb-final language, will any of the clitics be adjacent to the verb proper; since they are not bonded specifically to the verb, they cannot properly be considered part of "verb inflection". The only true personal inflection on the Hittite verb is the subject marker (distinct from the just-mentioned subject clitics). Old Lithuanian also has object clitics, which may occur either infixal

¹⁴ Breton apparently uses infixal pronouns exclusively.

(after preverbs) or suffixed (Senn 1966:191). These need not cliticize only to the verb, however,¹⁵ and moreover may occur as independent words; thus it would seem inappropriate to count them as part of verbal inflection. Albanian does have pronominal object markers which are true verbal clitics (Camaj 1984:94), and indeed both direct and indirect markers can cooccur. Note, however, that Old Lithuanian and Albanian are attested only very late. In Modern Lithuanian only a 3-person reflexive clitic form survives, a plausible indication that the clitic phenomenon as a whole does represent a (receding) archaism. In Albanian the phenomenon may instead be on a par diachronically with the late development of clitics in Romance, rather than reflecting an IE survival.

7) Infixing/suffixing alternation of pronominal object affixes

a) In Berber there is a well-defined block of verbal affixes (IndObj + DirObj + Deictic) which moves as a whole, positioning itself after simplex verbs (suffix) but between preverb and verb stem of complex verbs (infix):

fki- γ -as-t-iD

gave-I-him-it-hither

ur -as-t-iD fki- γ ara NEG-him-it-hither gave-I NEG

(respectively, "I gave it to him", "I did not give it to him": Chaker 1983:138-40, 150). The preverbs in question are strictly functional: they convey grammatical information (aspectual, negative, relative) and

¹⁵ Senn gives examples where the host word for the clitic is a clause-initial relative pronoun and a clause-initial conjunction.

are not part of the verb's lexical representation. The "participial" suffix -n shows the same mobility, but only after certain preverbs, notably NEG (Chaker 1983:119); here there is dialectal variation (Basset 1949). The form of the affixes is (apparently) identical whether infixed or suffixed. --- There is nothing analogous in Semitic or Egyptian.

b) In Old Irish, object pronouns are regularly suffixed to simplex verbs, infixed to complex verbs (Thurneysen 1946:255, 270):

ort- <u>i</u> "it killed him"	ro - <u>m</u> -gab "he took me"
it.killed-him	PERF-me-he.took

The process differs from Berber in a number of minor ways. First, suffixation only occurs in some environments. Suffixes are only attachable to certain person/number combinations of simplex verbs, and only in certain tenses (Thurneysen 1946:270-71); if these conditions fail to be met, the dummy preverb no- is attached to the verb and the pronoun assumes its infixed form. Second, even in suffixing environments, infixation (using no-) may sometimes occur anyway (271).¹⁶ Third, the infixed and suffixed forms of the pronouns are not identical; moreover, there are several distinct series of infixed forms (257ff.). Finally, the preverbs may be grammatical or lexical or both: Irish verbs (unlike Berber) are very often lexical complexes, "Preverb + Stem". Fairly elaborate rules govern the placement of the infixed pronoun in the event of multiple preverbs. --- Infixed pronouns also exist in Welsh; here, however, the preverb is always a grammatical one except in the earliest

¹⁶ Thurneysen asserts that infixation is always an option; Cowgill 1987 shows that this statement is much too strong.

poetry (Evans 1976:56), for the old inherited lexical preverbs became a fused and non-segmentable part of the verb stem early on.¹⁷ In general, an infixing/suffixing alternation is precluded by the almost total nonexistence of true suffixal pronouns [6b]; where these do exist (in early forms of the verb "to be"), one can indeed speak of such an alternation, since "to be" does normally take infixed pronouns like any verb (57). Breton, with no suffixed pronouns, has no alternation. Continental Gaulish has Obj markers as both infixes and suffixes (see [6b] above).

c) In Indo-European, both Albanian and Lithuanian show an alternation in the position of object markers. The Albanian phenomenon involves a suffixing/prefixing alternation: after imperatives the clitic is postverbal, otherwise preverbal (Camaj 1984:94). The Lithuanian object clitics occur either postverbally or in infix position (after preverbs); but the status of the markers as part of "verbal inflection" may be unclear. See [6c] above on this point, and also on the late date of attestation of these languages. --- None of the truly old IE languages marks objects on the verb at all, thus ruling out any positional alternation.

8) Position of definite article in Genitive embeddings: "house the-man"

a) The normal genitive technique in older Semitic is the "Construct",

¹⁷ Note the use in later Welsh of mi or fe as dummy preverbs, analogous to Old Irish no-.

whereby the two nouns in question are juxtaposed in the order Head-Dept = N-Gen; the head assumes special "Construct" morphology,

Hebrew: bayit (non-Construct) vs. bēt (Construct) "house", and the resultant N-Gen syntagm acts as a very tightly bonded, quasi-compound unit. The intimacy of this bond manifests itself on all levels: syntax (irreversible Head-Dept order; no adjective may intervene between N and Gen), morphology (special construct form or ending for the HeadN), and/or phonology (reduced and/or unaccented form of the HeadN). Some of the languages have noun case, some do not; the syntax of the Construct is essentially the same in either event.

Not all Semitic languages have a definite article; in those that do --- Hebrew ha-, Arabic (a)l-, Aramaic -ā (suffixed) --- the article is invariant. The only way to make the Construct syntagm definite is to attach the article to the Dependent, as in Hebrew:

bēt ha-ʔiṣ "house the-man" (= "the man's house")

This rule is extremely rigid, holding even in essentially lexicalized, compound-like combinations, e.g. Hebrew

qodeš ha-qədošim "holy the-holies" (= "the Holy of Holies")

On the other hand, Aramaic (not Biblical Hebrew and Arabic) has a competing, "periphrastic" genitive strategy using an independent "of" particle dī; here Head or Dependent or both may freely take the article, depending on the intended meaning:

šalīṭ - ā dī malk-ā "the king's captain" (Daniel 2:15)
captain-the of king-the

In Egyptian, which has both Construct-style and periphrastic genitives, definite articles develop only in Late Egyptian; here all patterns of

semantics.¹⁸ --- Continental Celtic apparently had no article, though the demonstrative element *sindos underlying the later article is attested. Falc'hun 1976 argues that an article an(d(e))- already existed as an initial element in Gaulish names. Indeed, the initial *s- of *sindos can delete in Gaulish.¹⁹ But Falc'hun's view has found little support; one might better interpret ande- as an element meaning "in", or as an intensifier.²⁰

c) Of the other old Indo-European languages, only Classical (post-Homeric) Greek and Old Norse have a clearly profiled definite article (distinct from demonstratives): Greek ho (proclitic), Norse -inn (usually suffixal). Nouns and articles in these languages are fully inflected for gender/number/case, and the orders N-Gen and Gen-N can both occur. There is no periphrastic construction with a separate "of" word. In Greek each noun takes its own article separately, and both articles appear in the genitive construction. Various orders are possible, including a "sandwiched" construction that breaks up article and noun:

¹⁸ In Welsh, the periphrastic construction seems restricted to nonpossessive uses of the genitive (torque of gold; each of them; swarm of bees). I would conjecture this to be the case elsewhere in Insular Celtic too.

¹⁹ Thus indas mnas "these women" from the Larzac inscription, but also in-sinde with the s preserved; see Lejeune et al. 1985:145, 152-53, 155, 168-69. On Gaulish demonstratives see Lejeune 1980.

²⁰ See Evans 1967:136-41. My thanks also to John Koch and to Stefan Schumacher.

ho [tou̇ anthrópou] oíkos "the house of the man"
 the [the.Gen man.Gen] house

Norse does appear to favor a Semitic-like restriction: when the order is specifically Head-Gen, the head tends not to take an article (Heusler 1950:125-26).²¹ Heusler presents examples such as the following (article is underlined):

Head-Gen: vald lands-ens "dominion of (= over) the land"
 Gen-Head: suerps hiołto-nom (Dat) "the sword's knob (Dat)"

The restriction may conceivably be broader and somewhat different from the Semitic one: Heusler does not say so, but his examples seem consonant with a generalization that the article (if any) may tend to follow the genitive syntagm as a whole, regardless of Head-Gen or Gen-Head order. Note that the use of the article in Norse is relatively sparse anyway (125), making a double-article construction (of the Greek sort) rather unlikely.²²

d) Comment: The category of definite article does not seem reconstructible to Proto-Semitic; not all the languages have an article, and in those that do, the morphemes attested are (probably) not all cognate. It is significant, then, that in the three Semitic languages where the category does appear (Hebrew, Arabic, Aramaic) the identical article-

²¹ "Artikkellos ist im allg. das Hauptwort vor Genitiv ... während das Hauptwort nach Genitiv auch Artikel annimmt" (125-26); see also examples on p. 177.

²² My thanks to John Lindow for discussion of this and other points of Norse grammar.

placement rule holds, with the same rigidity of application. Why the triple parallel evolution? In broad terms, the key is undoubtedly the interplay between the article rule and the Construct. As remarked, the Construct, which does reconstruct to Proto-Semitic, is itself characterized by extreme rigidity: the two linked nouns, locked together in a quasi-compound bond, are deprived of much of their autonomy. It seems natural, then, that in the Construct "definiteness" should apply not to the two nouns separately, but to the semi-fused N-N unit as a whole --- which is what in fact happens. And the target site for the article is that word in the Construct which takes the phonological stress: the last one, the Dependent. (The mismatch here between syntax and phonology, with the grammatical dependent behaving like a "phonological head", is a hallmark of Semitic.) To a considerable degree, then, the article placement rule has internal structural motivation in Semitic, and could well have arisen several times independently. (Significantly, the rule has no natural internal motivation in Celtic.)

In Celtic, the N-N genitive bond lacks the intrinsic tightness of the Semitic Construct: adjectives can intervene between the two nouns, the usual Head-Dept order can be reversed, especially in poetry (analyzed as compounding), and the bond is usually not marked phonologically (no mutation except under special conditions). Much more than in Semitic, then, the two nouns retain their autonomy; and as such they have a greater potentiality for separate definiteness marking. Hence the flexibility of the article-placement rule in Celtic vis-a-vis Semitic, and the un-Semitic possibility of two articles cooccurring.

Thus in both language groups, the article placement rule and the

N-Gen bond are correlated as regards "tightness": both phenomena are rigid in Semitic, loose in Celtic. It might appear that this difference between the two families should preclude any hypothetical substratal account; but the difficulty is illusory. On a substratal account, speakers of a Semitic-like language would have transferred their own speech habits to the syntax of imperfectly mastered Celtic. The Construct, so intimately intermeshed with Semitic-specific morphophonology, could not itself have been transplanted into Celtic. But a rule of article placement would not be cemented into the concrete stuff of Semitic grammar in the same way, and plausibly could "break loose" from its embedment in Semitic morphosyntax and be taken over as a phenomenon in its own right. And this rule, once divorced from the Construct, would be under no intrinsic structural compulsion to apply rigidly.

9) Nonconcord: verb before noun subject (V-Subj) takes maximally unmarked form

a) In Egyptian, in the environment V-Subj (with noun subject), the verb occurs as a bare stem with no person marking whatsoever (not even a 3rd-person suffix):

sdm-f "he hears" vs. sdm ntr "the god hears"

In the same environment in Arabic, both singular and plural subjects must take a singular verb (masculine if the subject is biologically male, feminine if female, usually feminine if inanimate):

kataba(3-m-sg.) t-tullāb(pl.) "the students wrote"

Full concord is standard for the marked Subj-V order in these languages (and in Akkadian, where Subj-(Obj)-V order is regular). --- Hebrew and

Ethiopic allow nonconcord (3-sg) with plural noun subject as an option, especially in V-Subj order. Concord remains the norm in Hebrew; in Ethiopic, nonhuman plural nouns may freely take singular or plural verbs, while human subjects generally show concord (Dillmann 1907:500-502, Lambdin 1978:27). --- The Berber verb always shows full agreement with its subject; "conjoined" singular NPs take their verb in the singular, but only because "N and N" is expressed as "N with N", so that the second "conjunct" is not actually in direct construction with the verb.

b) In Welsh of all periods, in V-Subj order (with noun subject) the verb standardly appears in the 3-sg form (though exceptions occur in Middle Welsh):

daeth / daethant	"he came" / "they came"
daeth y dynion	"the men came"

In the marked Subj-V order, full concord is the rule, again with exceptions (Evans 1976:179-80, Morris-Jones 1931:190-91). Breton (like Modern Irish) goes even further: in V-Subj order, the verb is 3-sg before an explicit subject of any sort, including non-3rd-person pronouns. And unlike Welsh, the Breton verb exhibits nonconcord even in Subj-V order (except in the negative):

huy a mano aman	"you will stay here"
you PTCL stay(3-sg) here	(Hemon 1984:273-74)

Nonconcord appears to be a very old phenomenon in Brythonic, occurring "as a predominating feature" in Cornish as well as Breton and Welsh (D. S. Evans 1971:49). --- Old Irish, however, observes full person/number agreement between verb and subject noun.

c) In no other older Indo-European language does the Brythonic/Egyptian/Arabic nonconcord pattern hold as the norm. Various subpatterns do exist where number concord may fail to hold. Ad sensum concord of verb with subject noun, for example, is not uncommon. In a number of languages (e.g. Greek, Old Norse), the verb may occasionally appear in the singular if the order is V-Subj, or if the subject is a conjoint of several singular nouns.²³ In Greek and Hittite, neuter plural subjects are construed as collectives and hence take singular verbs. In no instance, however, is nonconcord a clear rule as in Brythonic Celtic.

d) Comment: Note that two distinct phenomena are being conflated here: the total absence of any person/number marker (Egyptian), and the use of the 3-sg form in non-3-sg contexts (Brythonic).

Comment: It would seem natural that, in flexible non-VSO languages (such as most of IE), a salient environment for possible nonconcord should be V-Subj order --- as if, in such instances, the subject were not yet firmly fixed in the speaker's mind when the verb was uttered. (IE reference grammars sometimes explain the phenomenon in just this way, e.g. Heusler 1950:141 for Old Norse.) And for the same reason it seems just as natural that, in full-blown VSO languages, nonconcord should frequently apply as the standard agreement rule. The two

²³ The relevant Old Norse phenomena are not as clearcut as Wagner implies (1959:218), however. Verb-initial sentences are not unusual in Old Norse, but full concord is the rule; such sentences only show nonconcord when they convey an impersonal nuance (roughly the type "there appears two men..."). As for the "conjoined subject" type, it in fact splits the two conjuncts around the verb (the type "John came and Bill") (Heusler 1950:141).

phenomena (V-Subj order and nonconcord) would seem to correlate --- whether as option or as norm.²⁴ In light of this, the South American VSO language Yagua will provide a salutary lesson (see Appendix 2, and sec. 6.6.3.2): here nonconcord holds always and only when the order is Subj-V!

10) Verbal noun (VN: Obj in genitive) rather than infinitive (Obj in accusative)

a) The Semitic languages vary. Hebrew has a true infinitive, whose Obj (if definite) takes the accusative preposition et, as with finite verbs; Arabic has a verbal noun; Akkadian and Ethiopic have a VN/Infinitive which can take either nominal or verbal case rection, though nominal (genitival) rection appears commoner in both languages (von Soden 1969:202, Dillmann 1907:263).

Hebrew: $\text{lir?ot et ha-?i\text{š}}$ "to see ACC the-man"

Arabic: $\text{qawlu(nom) l-ḥaqqi(gen)}$ "to speak the-truth"

The Egyptian "infinitive" takes its object in the genitive (distinguishable from "accusative", in this caseless language, only for pronominal objects), and hence properly counts as a VN (Gardiner 1957:225). In Berber, all verbs have a corresponding "nom d'action" (Basset 1952:23). Functionally, however, these do not seem to belong to the verb's inflectional system, but rather have the status of derived nominals (not

²⁴ Note here Greenberg's "Universal" #33: "When number agreement between the noun and verb is suspended and the rule is based on order, the case is always one in which the verb precedes and the verb is in the singular" (1966a:94).

"singing" but "song", to take an English analogy); thus Chaker (1983:193-94) comments that "le plus souvent, seule la valeur de nom concret est conservée." Certainly the language makes heavy use of finite verbs (aorist) in prototypical "infinitival" contexts: "I want (that) he/I-go".

b) Insular Celtic has verbal nouns, with the object appearing as a normal genitive. In Brythonic this is clear despite the absence of noun case: for, with verbal nouns, pronominal objects take the form of possessive pronouns, while noun objects do not show the direct-object lenition found (in Welsh) after finite verbs. Thus Welsh:

gwelais fachgen	"I.saw a boy"	(<u>bachgen</u> lenites)
gweled bachgen	"seeing (of) a boy"	(= to see a boy)
ei weled	"his seeing"	(= to see him)

In Old Irish, which does have case, the genitive rection is transparent:

ni foílsitis	deícsin	a gnúsa	"they would not have endured
not endure.3pl	seeing	his face(gen.)	to see his face" (Wb. 15a20)

--- Another construction also exists in Irish, featuring the VerbNoun with preposition do "to" and a preceding object (Thurneysen 1946:445):

tol	dáe	do denum	"to.do God's will"
will(nom.)	God	to do	

In Modern Irish this construction [Obj do VerbNoun] is functionally an infinitive (see e.g. Fraser 1912:219). In Old Irish, however, the case marking on the preceding object NP apparently reflects, not its role in

the embedded "infinitival" clause, but rather the role of the embedded clause as a whole in the matrix clause;²⁵ thus the above example, in fuller context, translates as "that there be no care on anyone save doing God's will" (Dillon 1955:112), where a main-clause nominative is indeed appropriate (predicate of "be"). Following Disterheft's insightful analysis (1980:149-50, 154), I take this type as involving conceptually passive verbal nouns whose "object" (now passive subject) has undergone Raising into the matrix clause ("God's will to-be-done"). Note that on the criterion used here, this construction counts as neither infinitive nor verbal noun, for the object is no longer governed by the VerbNoun at all.

c) Outside of Celtic, all the older Indo-European languages except Vedic Sanskrit have clearly defined infinitives, governing an accusative object in the same way as the finite verb:

Latin: uidēre puerum(acc.) "to.see a boy"

In Vedic there are over a dozen morphological formations fulfilling roughly the function of an infinitive; some govern the accusative, some the genitive, some the dative (Delbrück 1888:410-25). --- It should be pointed out that "verbal nouns" (using the term in a peculiarly Indo-Europeanist sense) are richly attested throughout the family; see e.g. Rosén 1981 for Latin, and Porzig 1942 for Greek. As is clear from such monographs, however, these forms (unlike infinitives, and unlike the Semitic or Celtic verbal noun) are not really "inflectional". They do

²⁵ Thurneysen is less than explicit here, but the point emerges from his examples and from Fraser 1912 and Baudis 1913b:405ff., esp. 415; see further Dillon 1955:111-13.

not belong straightforwardly to grammar, but may occupy a variety of positions on a continuum between grammar and lexicon. They are often not fully productive, their semantics can be idiosyncratic, and to varying degrees they are less than fully integrated into verbal morphology either formally or functionally. (Recall the above discussion of Berber.) Such "verbal nouns" seem not greatly dissimilar to the formations in Vedic, just referred to.

d) Comment: As a cover term embracing both "verbal noun" and "infinitive", I will use the term "verbal abstract". Semantically, a verbal abstract must describe the verbal action per se; morphologically, it must be a regular part of the verbal inflectional paradigm.

Comment: Even languages which lack a case system usually have unambiguous ways of indicating "genitive" vs. "accusative" rection (see above), and it is in this extended sense that I use the traditional case labels. As we will see, some languages truly have no way of formally distinguishing the two.

Comment: I am (perhaps somewhat arbitrarily) taking the difference between genitive and accusative rection of the object as criterial for the concepts "verbal noun" and "infinitive". Traditionally Indo-Europeanists, in discussing the status of the infinitive, have also appealed to other criteria (see e.g. Meillet 1964:280-81, 375).²⁶ Thus, in addition to its accusative rection, the "ideal" infinitive as found

²⁶ For modern comparative treatment of the syntax of the infinitive in Indo-European, see Disterheft 1980. Inter alia, the book devotes considerable attention to a functional and syntactic characterization of what should constitute the notion "infinitive". Joseph 1983:6-36 also addresses the conceptual question of finiteness vs. infinitives, now in a Balkan context.

synchronically in the system of (say) Latin is a morphologically regular part of the system of verbal inflection: it is predictable in form (in most instances), does not itself inflect for case (like a noun), is built from the verb stem (as befits an inflectional category), and frequently codes voice and/or tense like a finite verb. This stands in contrast to the earlier situation seen in Vedic Sanskrit, whose various verbal nominals are not well integrated into the verb system, lack formal homogeneity, can inflect for case according to function in the matrix sentence, and are derived directly from the verb root. Such a multi-factor approach does not easily lend itself to crosslinguistic comparison. It is highly Indo-European-specific, and conflates many logically independent features which have not (to my knowledge) been shown to form a recurrent crosslinguistic cluster of properties. For purposes of typological comparison, one should ideally look at all these factors separately; provisionally, a single factor should be selected which epitomizes the nouniness/verbiness of the VN/Infinitive. Case rection of the governed object seemed best to capture this difference.

Comment: If the Vedic system is taken as representative of the situation in PIE, as indicated by the discussion in Meillet, then the regular development of the category "verbal noun" in Celtic (with object in genitive) is just as marked a departure from the PIE system as is the regular development of infinitives elsewhere in Indo-European (with object in accusative). In both instances, original heterogeneity of object rection has been smoothed out into a functionally regular system --- but in two different directions. (Vedic and Irish do resemble each other in building their verbal nouns on a rich variety of morphological patterns, and in the fact that the verbal noun does show case inflection

like an ordinary noun. Arabic agrees with Vedic and Irish in both these respects.)

11) Predicative particle identical (or homophonous) to a "local" preposition

a) In Egyptian, sentences rendering the meaning "Subj is NominalPred" may do so with a predicative particle m which is spelled the same as the preposition "in" (Gardiner 1957:40-41):

(BE) Subj	<u>m</u>	NomPred	"Subj is NomPred"
iw -f	m	smr	"he is a companion"
is -he		companion	

There is, additionally, a future-tense predicative particle r appearing in the identical syntactic frame, homographic with the directional preposition "to" (Gardiner 1957:97):

(BE) Subj	<u>r</u>	NomPred	"Subj will-be NomPred"
iw -f	r	smr	"he will be a companion"

The double semantic tense correspondence --- "in" : present (actually zero/unmarked), "to" : future --- is intuitively natural, and argues strongly that the predicative particle actually is the preposition in both instances (rather than a mere homograph or homophone). Note that Egyptian also has various other possibilities for noun and adjective predicates, usually featuring the order

NomPred Subj

and taking neither copula nor predicative particle. --- In Berber, similarly, nominal predicates are preceded by a predicative particle d that is homophonous with the preposition d "with" (Chaker 1983:321):

d NomPred (BE) Subj
 d amak^war wrgaz "the man is a thief"
 thief man

Here the two elements d are less obviously "the same", for Berber has several other particles d: a tense/aspect particle, a relative particle (in some dialects only), and a proximal deictic particle "here/hither". A priori the predicative d might be linked to any of these, and/or to "with". --- In Semitic, predicate nominals usually take no particle; in case-marking Arabic, the case of the NomPred is nominative in zero-copula sentences, but accusative after the tensed copular verb kāna. There is, however, a marked variant in Arabic where the preposition bi "in" may occur before the predicate nominal, notably in negative sentences; very rare traces of the same thing exist in Hebrew (but not, to my knowledge, in Akkadian or Ethiopic). Plausibly this represents a receding archaism in Semitic (Gensler 1991).

b) In Welsh, a similar rule holds (Evans 1976:139): a nominal predicate, when it follows the subject, may be preceded (not obligatorily) by a predicative particle yn which is homophonous with the preposition "in":

BE Subj yn NomPred
 y mae ef yn ffermwr "he is a farmer"
 is he farmer

These various yns, however, differ in the mutations they trigger: nasalization following the true preposition, lenition after the predicative particle (and no mutation when yn is used with the verbal noun in periphrastic verb tenses; see [12b] below). In all likelihood predicative

yn is of different diachronic origin than the two other yns (Evans 1976:216, apud T. Arwyn Watkins [1957, 1960, 1962]); in Old Welsh, predicative yn was sometimes spelled int, plausibly a reflex of the old Celtic article *sindos. --- Note that Welsh also has other predicate-nominal patterns, in which the predicate nominal precedes the subject; here yn does not appear, and the copula (in older language) is optional (Evans 1976:139-40):

- (i) (BE) NomPred Subj or (ii) NomPred BE Subj.

Construction (ii) is a normal option in Modern Welsh, alongside the construction with yn. --- Breton has no predicative particle; the normal construction (Hemon 1984:262-63) is

NomPred BE Subj

maro eo ma chatal "my cattle are dead".

Nor is there such a particle in Old Irish; the normal pattern is

BE NomPred Subj

it cóem ind eich "the horses are precious"

are precious the horses (Stories from the Táin, 13),

with the copula BE procliticized to the predicate (Thurneysen 1946:475).

Both of these resemble the Welsh construction without yn. Modern Irish does have the syntagm "he is in his doctor", meaning "he is a doctor" (note that the substantive verb is used, not the copula); but this is a late phenomenon (Dillon 1927-28:324-26) and thus not at issue here.

c) Nominal predicates in most older Indo-European languages appear in the nominative case and take no particle. In many Balto-Slavic languages, nominal predicates regularly take the instrumental case (on this see Nichols 1973); the phenomenon has been on the rise in Slavic

since the earliest documents (1973:2). The same thing crops up occasionally in Armenian and Sanskrit (1973:20). This occurrence of a "local" case is weakly analogous to the occurrence of "local" prepositions as predicative particles in the syntagm at issue here. (In Modern Russian, the instrumental occurs only when an explicit copular verb is present, the nominative appearing in pure nominal sentences.) None of the older languages, however, makes use of a predicative particle. In one modern Slavic language --- Sorbian --- the inherited "bare" instrumental has been replaced by the preposition "with" + instrumental (1973:201). The result is a perfect typological parallel to the Welsh/Berber/Egyptian type, though unconnected with any of these languages diachronically.

d) Comment: Even if the two Welsh elements yn have different etymologies, this need not ipso facto exclude the possibility of hypothetical substratal influence. Bilingual speakers whose native (substratal) language had homophonous or identical Pred/Prep morphemes might well have transferred this speech habit to pre-Welsh, fostering a convergence of two similar but originally distinct Brythonic morphemes.

Comment: Though hardly a universal generalization (Berber), it is noteworthy that in three languages --- Arabic accusative after kāna, Welsh yn, and Russian predicate instrumental --- the special predicate mark (particle or case) occurs only when an explicit copula is present.

Comment: In the CHS languages there is a close link between this feature and feature [12], "prepositional periphrastic tenses"; see [12d] below.

12) Prepositional periphrastic tenses: "BE Prep+VN"

a) Very common in Egyptian, with prepositions hr, m, r ("on, in, to", cf. [11a] above); as in [11a], r "to" adds an explicit future nuance otherwise absent. In the present (or unmarked) tense, the usual particle is hr "on", though m "in" is also possible --- a minor difference vis-a-vis the construction in [11], where hr does not occur. Thus:

iw	srw	hr	rdit	n-k	"the nobles give to you"
BE nobles on giving to-you					(Gardiner 1957:247)

The construction has no currency in Semitic or (as far as I know) Berber.

b) In Welsh the construction is very common, involving an element yn which is homophonous to the preposition "in", though in this construction yn is followed by the unmutated form (cf. [11b] above); the VN can be fronted, in which case yn is optional (Evans 1976:138). Thus:

y	mae	ef	yn	canu	"he is singing" (ModW)
is	he	singing			

The same construction can occur with the preposition wedi "after" to express past time.²⁷ As with Egyptian ([11a, 12a]), the naturalness of the double semantic tense correspondence --- "in" : present, "after" : past --- argues strongly for a synchronic identification of the particle yn with the preposition (rather than being an accidental homophone). In older Breton the prepositional periphrastic makes use of the preposition

²⁷ In Modern Welsh, the preposition ar "on" in the same construction expresses near futurity ("about to, on the point of"); in Middle Welsh, ar can occasionally occur with verbal nouns, but with present semantics (Evans 1976:186).

ouzh "at, against";²⁸ this later becomes o(c'h), a form no longer homophonous to a preposition (Hemon 1984:268-69). Breton characteristically fronts the Prep+VN:

o labourat en e bark emañ	"He is working in his field"
at working in his field he.is	(Desbordes 1983:70)

In Old Irish the construction (with preposition oc "at"), though not yet the all-pervasive phenomenon of Modern Irish, is not uncommon even in the glosses (Wagner calls it "recht geläufig" [1959:126, cf. 239]); e.g.

in tain nombíu oc irbáig air-ib	"when I am glorying
the time I.am.REL at glorying for-you	for you" (Wb. 20a3)

c) Not found elsewhere in older Indo-European; attested at later periods in (e.g.) Middle English, "BE on VERBing".

d) Comment: A word needs to be said about the distinction between identity and homophony of particle and preposition in features [11], [12]. In those CHS languages which have both constructions, it is natural to examine the two together:²⁹

[11] BE SUBJ Ptcl + NomPred

[12] BE SUBJ Ptcl + VN

²⁸ In Modern Welsh, the cognate wrth "to, by" can likewise appear in prepositional periphrastics, in the meaning "be engaged in doing"; for Middle Welsh I found no indication of this (Evans 1976:213-14).

²⁹ The question of whether features [11] and [12] represent a natural coupling universally is another matter entirely, which will be addressed in sec. 6.6.3.3.

The two constructions do cooccur in Egyptian and in Welsh, while Breton and Irish have the latter alone. In Egyptian, the identification of the particle with a preposition (in both constructions) is not problematical; the "tense semantics" argument of [11a] above (r = future) applies equally well here ([12a]), and the very fact that it applies twice strengthens the force of the argument. Functionally, too, a prepositional analysis makes sense in both [11] and [12], and in analogous ways: a nominal predicate and a (notional) participial predicate, respectively, are being recast as an adverbial predicate, realized as a PrepPhrase (see Gardiner's discussion, 1957:244). In Breton and Irish, only construction [12] occurs; but here again the identification of the particle as a Prep is straightforward (in particular, the particle triggers the same mutation as would the preposition). Structurally, then, Egyptian, Breton, and Irish follow essentially the same pattern, here realized most fully in Egyptian.

Only in Welsh are there objections to a prepositional analysis. Synchronically, each of the three uses of yn shows a different mutation: lenition in [11], nonmutation in [12], nasalization after the ordinary preposition. Diachronically, the predicative particle arguably did not originate in the preposition, being spelled differently in Old Welsh (see discussion in [11b]). But how much weight should be given to these objections? The diachronic problem need not affect the Middle Welsh synchronic analysis; as argued in [11d] above, there is no reason that originally distinct particles could not have converged. The problem with the mutations is valid, but its force is weakened by the fact that some word-initial segments (vowels, some consonants) do not mutate at all. Counterbalancing this is the "tense semantics" argument (wedi, yn)

--- applicable in Welsh (unlike Egyptian) only to construction [12] --- which provides strong structural support for the equation $Ptcl = Prep$. Further, regardless of the status of Welsh, the "prepositional" pattern clearly represents a natural structural configuration both within Insular Celtic (witness Irish, Breton) and more broadly (Egyptian). It is instructive, too, that the Irish and Breton prepositional periphrastics are not tied to any one specific etymon but involve two unrelated prepositions (Irish oc, Breton ouzh; see Thurneysen 1946:515, 525), with Welsh yn making a potential third; the three constructions thus would have been at least partially independent of one another, testifying further to their structural naturalness in Celtic. It seems to me, then, that within these languages an equation $Ptcl = Prep$ (tempered somewhat by the difference in mutations) makes far better sense as a synchronic analysis of Welsh (Middle as well as Modern) than an analysis which emphasizes the distinctness of the various elements yn.³⁰ The claim applies most securely to feature [12], where commonalities with other Celtic languages buttress the argument; for [11], no such intra-Celtic comparative justification is possible, and structural parallels with Egyptian must be appealed to.

13) "DO" periphrastic tenses: "DO + VN"

a) Common in Egyptian, and especially Late Egyptian, notably with verbs of motion and with "heavy" roots (Gardiner 1957:395). Example:

³⁰ For a very similar argument applied to Modern Welsh, see Fife 1990, especially pp. 310-16, 368-86, 422-42.

irt-i ^vsmt m hntyt "I departed (= made departure)
 DOing-my departing in south southwards"

(The example also illustrates the narrative use of the "infinitive" irt, see [15a].) Not found in Semitic or (as far as I know) Berber.

b) Very common in Brythonic, occurring in Welsh (especially Middle Welsh prose) and Breton (often to lend emphasis to the verbal action, Hemon 1984:249-50). In both languages, the VN is normally fronted. Thus Middle Welsh:

mynet a oruc Padric y Iwerdon "Patrick went to Ireland"
 going PTCL did to Ireland (Evans 1976:160)

The type is not characteristic of Irish.

c) Not found elsewhere in old Indo-European.³¹ A superficially identical construction is well-attested at a later period in Modern Persian, where it represents the normal construction with borrowed (Arabic) verbs.

14) Adverbial clauses of the form "and + finite clause"

a) In Hebrew and Arabic, adverb clauses of accompanying circumstance are very commonly formed from a nominal clause preceded by "and"

³¹ According to one theory, the Germanic weak preterite may have arisen from a fusion of verb-stem + "DO"; but this is a diachronic reconstruction, not a description of any attested synchronic reality.

(Gesenius 1910:453, 489). The nominal clause can lack a verb entirely, or the verb may appear in a nonfinite form. Thus:

va-yəxasu et ^cervat avi-hem ū-fnē -hem aḥoranīt
and-they.covered ACC nakedness father-their and-face-their backward
"And they covered the nakedness of their father, with their
faces (being turned) backward" (Gen 9:23) [Hebrew]

va- yēra ēl-av YHWH ... və-hū yošēv petah ha-ʔohel
and-appeared to-him God and-he sitting door the tent
"And the Lord appeared to him ... as he sat in the tent door"
(Gen 18:1) [Hebrew]

qāma zaydun wa -huwa bākin "Zaid arose weeping" [Arabic]
arose Zaid and-he weeping (Wright 1967:II 330)

This is the well-known construction known as ḥāl ("state, condition") in Arabic grammatical terminology. However, a finite verb can equally well appear after "and" in such constructions (Gesenius 1910:456, 489; Wright 1967:II 330-33), normally with inverted word order in the "and"-clause (Subj-V, not the normal V-Subj):

al -təʔaḥarū otī va -YHWH hitslīah dark-ī
NEG-you.delay me and-God he.prospered way-my
"Do not delay me, seeing as God has prospered my way"
(Gen 24:56) [Hebrew]

ka^xabtu^m wa -ʔantu^m ta^clamūna "you lied knowingly" [Arabic]

you.lied and- you you.know (Wright II 330)

ka^xabtu^m wa -ʔantu^m ta^clamūna

The same construction with "and" exists in Ethiopic, but it is "not nearly so common in Ethiopic as in Arabic" (Dillmann 1907:523). Akkadian also makes vigorous use of sentence coordination to express notional (adverbial) subordination, often with the two clauses separated by the common enclitic -ma "and so" (von Soden 1969:210-11, cf. 177-78). --- Neither Egyptian nor Berber has a conjunction "and"; both languages can link two nouns with "with", but this device cannot be used to conjoin two clauses. Hence the construction does not exist in these languages. (Wagner 1959:214 does cite a Berber example, with the Arabic borrowing u serving as the conjunction "and".)

b) Both Welsh and Irish very commonly form clauses of accompanying circumstance with "and" + nominal clause. Here, however, the restriction to nominal clauses is strict; a finite verb may not appear. As with Semitic, a verb may be lacking entirely, or it may appear in a non-finite form. Thus in Old Irish:

téit ass íarum, ocus a sciath slissen laiss

he.goes out then and his shield chip(gen.) with.him

"Then he goes out, his wooden shield (being) with him"

(Stories from the Táin, 2)

dobertis cech n-olc ... form os mese oc taircitul
 they.brought every evil on.me and I at prophesying
 cech maith doibsom
 every good to.them

"They inflicted every evil on me, while I was prophesying
 every good to them" (Ml. 54c30)

The Celtic and Semitic constructions are almost exact parallels, with the Semitic participle corresponding (unproblematically) to the Celtic "Prep+VN".

c) For some of the older Indo-European languages the grammars speak of parataxis, in the sense of "the arrangement of two independent sentences side by side, though one is in thought subordinate to the other" (Smyth 1956:485, for Greek). This certainly exists in IE; but the construction typically occurs both syndetic and asyndetic, with the conjunction "and" playing no essential role. --- Wagner (1959:213-14) presents the verb-less "and" construction as occurring in Old Norse, citing Heusler 1950:184, and in Old English.³²

d) Comment: The identifying characteristic of this type is the presence of an explicit conjunction "and", followed by a finite clause; thus notional subordination is achieved formally through pure coordination.

³² Note, however, that Heusler characterizes the Norse construction as "ausseralltäglich" and presents it under the heading of ellipsis; his examples frequently involve absence of the helping verb in compound tenses, with the participle remaining.

15) VerbNoun/Infinitive used instead of finite main-clause verb

a) In Hebrew, the so-called "infinitive absolute" can substitute for finite tenses (Gesenius 1910:345-47), sometimes to carry forward a narrative segment initiated by a finite verb (1st example below), sometimes in "bare" context (2nd example); thus:

və-^Cal har s̄inai yaradta və-^Cdaber ^Cima-hem
and-on mount Sinai you.descended and-to.speak with-them

"And you came down upon Mt. Sinai and spoke with them" (Neh 9:13)

haro ^Camal və-^Cyalod aven ū -viṭn -am
to.conceive mischief and-to.bear trouble and-belly-their
taxīn mirma
prepares deceit

"They conceive mischief and bring forth trouble, and
their belly prepares deceit" (Job 15:35)

Similarly in Phoenician. The construction apparently occurs in traces in later Akkadian (in a single inscription; see von Soden 1969:204). It is not, to my knowledge, found in Arabic or Ethiopic. --- In Egyptian, the infinitive can substitute for the finite verb in headings or book subtitles, or "to announce incidents of outstanding importance" (Gardiner 1957:229-30):

rdit-f wi m-h?t hrdw-f
placing-his me before children-his
"He placed me in front of his children"

(See [13a] for another Egyptian example.) For Berber, see the reservations in [10a].

b) In Middle Welsh the verbal noun often substitutes for the finite verb in narrative (Evans 1976:161), notably (though not necessarily) to continue the action initiated by a preceding finite verb; e.g.:

y dygyuores uyg kyuoeth ... a rodi dewis im
 PTCL rose my people and giving(v.n.) choice to.me
 "My people arose ... and gave me a choice"

The same phenomenon occurs in Breton (Hemon 1984:266-67). Wagner (1959:207-8) documents the phenomenon in Irish, but admits that it is nowhere near as prominent as in Welsh. On the other hand, the verbal noun does often appear in the Irish Annals in a "header" usage similar to that found in Egyptian (see e.g. Baudis^v 1913b:381); e.g.

indredh Mide la Niall "the invasion of Meath by Niall"

c) Predicative use of a VN/Infinitive would appear to be quite unusual in Indo-European. Wagner (1959:208) points out parallels in Old Norse, citing Heusler 1950:189-90. In Latin the construction is characteristic of certain "archaizing historians" such as Sallust and Tacitus (Disterheft 1980:196), though not of the early Latin dramatists; Disterheft mentions no similar uses elsewhere in Indo-European. Holland states that narrative infinitives "are found in Sanskrit, Greek, Latin, Old Norse, and Hittite" (1982:168); however, the remark (a passing footnote in an article on a different subject) says nothing about how normal or rare the phenomenon is. --- Indo-European does have a rough

functional analogue in the common "absolute construction", exemplified by the English type "John having arrived, I left". However, here the nonfinite form is not a VN/Infinitive but a participle; further, as in the English example, the nonfinite form almost always precedes the finite, not the other way round.³³

d) Comment: Two quite distinct usages are being conflated here: (1) the narrative continuation of a preceding finite verb; (2) the header usage.

Comment: The narrative usage standardly goes under the name of clause chaining; it is most common in its mirror-image version, in OV languages, with a string of nonfinite forms preceding and leading up to a finite form.

16) Syntactically governed word-initial change

a) Not found in Semitic or Egyptian. A fundamental feature of the grammar of Berber, where it is present as the alternation between "état libre" and "état d'annexion" (Free vs. Annexation state). Formally, the change does not involve consonantal mutation per se, but rather various initial alternations between vowels (Free:Annexation = a:u-) or between vowel and glide+vowel (V:GV = u:wu-, i:yi-) or in the feminine between tV and t (ti:t-, ta:t-) (see Chaker 1983:84, 92ff.). The choice among these alternatives is conditioned lexically. Examples:

³³ My thanks to Gary Holland for clarifying this and other points.

Free: argaz	Annexation: wrgaz	"man"
tamɣart	tmɣart	"town"

Functionally, each of the two variants covers a range of disparate uses, with Free state used for citation form, clause-initial topic, adnominal adjective, direct object, and predicate nominal, and Annexation state in most other environments (notably subject, genitive, object of most prepositions, and apposition to proleptic pronoun).

b) Found throughout Insular Celtic as "initial consonant mutation", though differing from language to language in actual realization (number and phonetic nature of the different mutations; functions served by the particular mutations). All the languages have at least "soft mutation", also termed "lenition" (fricativization in Irish, increase in sonority in Brythonic), in addition to the plain (unmutated) form. Thus with the Welsh word "father":

<u>Plain</u>	<u>Lenited</u>	<u>Nasal</u>	<u>Aspirate</u>
tad	dad	nhad	thad

Each of the mutations serves a very broad range of mixed functions, even more diverse than in Berber and varying from language to language. In Welsh, for example, the lenited form serves to mark (inter alia) the adjective (after feminine nouns), direct object, predicate nominal, and object of most prepositions; while citation form and clause-initial topic take the plain form, as do subject and genitive except under well-defined conditions.

c) Unknown elsewhere in Indo-European.

d) Comment: In purely formal terms, the Insular Celtic initial mutations arose through a metanalysis of original Indo-European grammatical endings; the mutations are the later reflexes of these endings, appearing word-initially on the word following the original Indo-European host of the ending. This formal account, however, does not go beyond providing a reasonable path of development; it does not explain why Celtic and only Celtic should have chosen precisely that path.

Comment: Berber and Insular Celtic do not show a point-for-point match in the particular functions served by their "mutated forms"; indeed, to a large degree the behaviors in the two languages are opposite (see brief list of functions above, and Gensler 1992).

Comment: It is not too uncommon to find languages where syntactically controlled word-initial change affects a single grammatical category of the language. Paths of phonetic change leading to such a phenomenon are not particularly exotic --- such developments as absorption of a prenominal or preverbal particle into the noun or verb. Examples are person (Tubu), number (Fula), causativity (Burmese), and realis/irrealis (Oceanic languages, per Lynch 1975). Note that these languages are not restricted areally but come from all over the world. What is remarkable about the Celtic and Berber phenomenon, by contrast, is its multicategorical nature, defying any easy form-function equation. Only Gilyak, of the non-CHS languages in my sample, has such a multicategorical phenomenon; it occurs also in Southwestern Mande.

Comment: It is likely that the Berber opposition is not an ancient feature of the language: the word-initial element (ti/t-, etc.) which carries the opposition of "free vs. annexation" states was probably once an article, a category which modern Berber lacks (see Vycichl 1957 and

Rössler 1960:145-46; also Greenberg 1978 for a general study of the life-cycle of articles). Some of the evidence for this claim dates from historical times (Vycichl 1957:139-140). For example, some old noun borrowings from Berber into dialect Arabic cannot take the Arabic definite article (a)l-; the role of the missing article, plausibly, is filled by the word-initial element of the borrowed Berber lexeme. And certain place names appear both with and without the initial element in Canary Islands Berber and in medieval Arabic historical writings. If valid, such considerations argue that Berber word-initial change as we know it did not come into existence at least until the time of the Arab conquest --- much too late to be implicated in any hypothetical pre-Celtic substratum on the British Isles.

17) Abstract or metaphorical use of kin terms: "son/father of Noun"

a) Common in Semitic, especially Arabic but also elsewhere --- e.g. Hebrew ben mavet "son of death" = one deserving of death, bat ^cayin "daughter of eye" = pupil; Arabic ibn al-sabil "son of the road" = traveler; Akkadian mār šipri "son of sending" = messenger; Ethiopic aba dam "father of blood" = blood avenger. The usage appears rather less common in Ethiopic and Akkadian. Adams also presents examples from Berber (1975:242). Not a feature of Egyptian. Semitic and Egyptian also make similar use of nouns meaning "possessor", but this seems semantically more natural and less idiosyncratic than the specific appeal to a kin term.

b) Common in Old Irish (e.g. mac báis "son of death" = wicked man, mac

imblissen "son of iris" = pupil of the eye, mac meda "son of mead" = drunkard, mac tîre "son of land" = wolf). On these see Adams 1975:240-42, who gives a long list. Not a prominent feature of Welsh.

c) Not found elsewhere in Indo-European.

d) Comment: The essence of this feature is its specific use of kinship terms in a true genitive construction (that is, not in a compound), with special nonliteral semantics: "KIN of Noun", in the meaning "person/thing characterized by some essential connection with Noun".

Comment: This feature stands out from the others in being not grammatical but lexical; it resembles them in appealing to structural considerations, in this instance the structure of the lexicon.

It should be noted that certain "un-Indo-European" features of Insular Celtic are not being examined here; for example:

a) The absence of the functional category of active participle in Celtic cannot be explained by appeal to Hamito-Semitic, since Semitic and Egyptian do have an active participle; the Berber "participle", used only in subject relative clauses, is not fully participial in character, but does embody one central function of active participles.

b) Impersonal verb constructions, found increasingly in modern Insular Celtic languages, are not a feature of Hamito-Semitic.

On the other hand, the above list is certainly not exhaustive. To mention only a single point, the morphological multiplicity of noun-plural formations found in (e.g.) Welsh and Arabic would certainly bear crosslinguistic investigation.

Chapter 4: Sources and sample

4.1 Orientation

With this chapter the investigation changes course. As stated in Chapter 1, our purpose will be to examine the Celtic/Hamito-Semitic (CHS) features in languages all over the world, and thereby gain some idea of whether the set of resemblances can or cannot plausibly be taken as "coincidence". The second half of this study is thus devoted to carrying out a concrete project of typological research. Chapter 4 covers general questions of sources and sample; Chapter 5 presents, in close detail, the scoring system whereby the rich variety of phenomena found worldwide was reduced to a set of numbers indicating closeness of fit to the CHS type; Chapter 6 outlines the results of the survey; Chapter 7 discusses the implications of the study; and Appendix 2 gives capsule data summaries for all the non-CHS languages included in the study, the raw material from which the scores emerged.

The shift in subject matter will carry with it a shift in the terms of the discourse. Those languages and constructs that have figured so prominently in all previous work on the problem --- Celtic, Indo-European, Hamito-Semitic, comparative grammar, diachrony, substrata --- will be relegated to the background. Their place will be taken by typological concerns: sample, features, feature splitting, feature bivalency, quantification, scoring, bias, ergativity, clause chaining.

Even in Chapter 6, where we take the results of the survey and reintegrate them into the original question, the typological way of thinking will be evident. And the CHS languages themselves recede into the background. Seen in typological perspective, their role is simply to define a set of features for global comparison. With that set now in hand (subject to modification), it will no longer be appropriate to pay special attention to these languages. Only when a feature must be rethought from the ground up, as will happen several times, will we again turn to the CHS phenomenon, as providing a prototype to which the revised feature must remain true.

4.2 Grammar books and the problem of silence

To a particularly acute degree, the present investigation was sensitive to the quality, scope, and even design of available grammar books. Given the sheer quantity and complexity of the features under discussion, only a grammar with a rich treatment of syntax could possibly provide the needed information. But the reasons ran much deeper than that.

To look at one language through lenses designed for another --- to ask CHS-oriented questions of languages in New Guinea and South America --- is an odd thing. In doing so, I was making unusual demands on the grammar books I used, asking questions which they were often not set up to answer.¹ For one thing, it is rare for a grammar book to state what

¹ The Croom-Helm series of grammars, with their explicitly panglottal orientation, proved a happy exception to this generalization.

its language is not like. There is normally no reason to do so, and one very good reason not to: the number of nonoccurrent features is limitless. A nonoccurrent feature will not normally be mentioned unless it represents an expected possibility, a structure known in advance to be characteristic of the family, area, or type to which the language belongs. Yet half the time, in looking at a given language, my goal was precisely to come to a definite determination about nonoccurrence, and to do so for CHS-style structural features which were not just nonoccurrent but alien to the language's whole structural cut. With rare exceptions, the grammar book's only answer to such questions ("Does Dyirbal have word-initial change?") was silence. I was thus repeatedly faced with the problem attendant on any "argumentum ex silentio": what does the silence betoken, nonoccurrence in the language or mere accidental omission from the source?² The larger and better the grammar book, the less the chance of accidental omission of an occurrent feature.

Even a phenomenon that does occur in a language can go undescribed, however --- even in a very good grammar book --- if the feature is not a clear and highly profiled aspect of the grammar. Sometimes the feature may seem so ordinary as to be invisible. In the CHS languages, for example, where articles in genitival embeddings are subject to a special positional constraint ("house the-boy"), the grammar books quite naturally have something to say about article placement in N-N embeddings. In languages without such a restriction, however, it would seem pointless to discuss article placement at all (though the phenomenon per se

² In this case, the answer is unproblematical --- a dramatic feature like initial change would hardly go unmentioned in a section on morphophonemics --- but things were not always so clearcut.

is perfectly real): typically the two nouns may each take an article, and a grammar writer would have to be clairvoyant to even think to mention the point.³ Here one can only infer the rule from examples. And the fuller the grammar, the greater the probable richness of the examples given.

A more subtle problem arises because similar etic facts can take on a radically different emic cast in different languages. What is a clear and highly profiled part of one language's grammar --- e.g. Prepositional periphrastic tenses (Prep + VN) in Celtic --- may exist in another language only as a minor configuration, perhaps catalogued in an unexpected part of the grammar book or presented as an incidental variation on a totally different syntactic theme. (This was indeed sometimes the case with Adpositional periphrastic tenses.) A relatively sketchy grammar may not bother to go into full detail on such minor configurations, especially if they fall into uninteresting or out-of-the-way corners of the description, and thereby may fail to mention precisely the information required by such studies as the present one.

Such problems with grammatical descriptions can arise in any global-scale typological study, of course. Many (most?) such studies, however, have dealt with features which are acknowledged to have universal significance in general linguistics and for which data tends to be straightforwardly available even in run-of-the-mill grammar books --- word order, person/number/gender markers, case marking systems,

³ Such things can happen even in superb grammar books. The grammar of Slave, for example, nowhere (?) addresses the syntax of copular sentences as such, though the language does have copular verbs.

head/dependent marking, valence-changing devices. Such features, in a sense, "belong" not to any specific language or languages but to the world and to language as a whole. The present study, in large measure, is different; perhaps half the features surveyed are exotics, more or less unfamiliar outside the world of Celtic and Hamito-Semitic grammar. The difficulties outlined above are thus particularly acute. To a great degree, therefore, my choice of languages was dictated by the quality of the grammar book, occasionally even at the expense of genetic and areal balance.⁴

4.3 Commensurability, reinterpretation, and hedged categories

To what extent should information gleaned from grammar books be taken at face value in a crosslinguistic survey? Quite apart from instances of obvious error or omission, any grammar is subject to problems of incommensurability with other grammars: the selfsame term will very often signify different things in different languages and in different linguistic descriptions. In an ideal world, the information would never simply be accepted as is, but would be subjected to a

⁴ Even an excellent grammar book cannot guarantee against mistakes on the part of the grammar user. Given the scope of the questions being asked, I seldom had the luxury of simply dipping into a grammar book for individual answers, but typically had to come to terms with the language as a whole, roughly yet accurately, in only a few days' time. Obviously the possibilities for error here are manifold: overlooking crucial information, misunderstanding what was said, misrepresenting the data in reinterpreting it. The scoring system for the data seeks to mitigate the effects of such error, especially judgment error, by building several kinds of hedged scores (half, zero) into the system instead of insisting on yes/no answers. Where possible, I double-checked with experts on the language. Some error of fact and interpretation undoubtedly remains.

critical reinterpretation. Such reinterpretation, however, would magnify an already large task enormously. It would also be an open-ended project; the ultimate goal, to make all linguistic descriptions truly commensurate, is as utopian as the comprehensive theory of language it presupposes. Throughout this study, the unmarked assumption has been that the information contained in grammar books can be accepted as trustworthy, and that categories are not being perversely redefined in radically unconventional ways.

In certain cases, however, I could not avoid a critical reevaluation. This was most apparent with regard to the distinction between infinitives and verbal nouns, one of the areas where the descriptive terminology is least standardized. Here I had to impose a terminological choice of my own, singling out the factor of genitive vs. accusative behavior of the notional object as criterial, and accordingly I took the terminological labels used in grammar books only as rough heuristics. But the identical difficulty also resurfaced one "metalevel" higher, with regard to the distinction between finite and nonfinite forms of the verb --- a distinction logically prior to that between infinitives and verbal nouns, both of which presuppose nonfiniteness. Here the terminological inconsistency was just as great, and again I often ignored the book's label (e.g. the term "nonfinite" in Abkhaz). Indeed, "nonfinite" has no single meaning in general linguistics; given the heterogeneous range of forms presented as nonfinite in grammars of the world's languages, any attempt at imposing a consistent crosslinguistic criterion must inevitably fly in the face of many language-specific descriptions. This is so in large measure because two competing and very different criteria are commonly appealed to: a form may be called

"nonfinite" by virtue of failure to code tense/aspect, or failure to code subject. Only in some languages are the two criteria congruent; in others they pick out completely different sets of verb forms. In some languages (e.g. Indo-European or Bantu) the subject-coding criterion would seem to better reflect the language's structure; in others (e.g. Yuman and probably other North American languages), the tense/aspect criterion would appear to fit better. Yet typological consistency requires that one of the two be systematically privileged at the expense of the other.

The take-off point for the present study is a set of phenomena defined relative to Indo-European and Mediterranean Hamito-Semitic. Here and throughout this study, I have allowed these languages to dictate terms to the analysis when conceptual difficulties arose. In this instance the difficulty concerns the problematical concept "nonfiniteness", for which conflicting criteria apply; and it appeared reasonable that such conflicts be resolved systematically in favor of IE and HS languages rather than (say) the languages of North America. The notion of "nonfinite" adopted should resonate with the IE and HS phenomena, rather than forcing them into strange categorial niches. Thus it should capture the common denominator uniting IE-style infinitives and CHS-style verbal nouns (feature [10]); the "nonfinite" forms it defines must be the kind of forms that can be thought of as embeddable CHS-style under Adpositions (feature [12]), or in combination with a verb like "DO" (feature [13]).

Accordingly, it is the absence or optionality of subject coding (over against obligatory subject coding on the finite verb) which I

chose as the principal criterion for "nonfinite". Such a criterion allows infinitives, for example, to have tense (as indeed they do in IE), while expecting infinitives and verbal nouns to be predominantly subjectless (again as in IE and HS). When applied to constructions where "nonfinite" forms are subordinated to Adpositions or to "DO", it will admit only subjectless forms --- in keeping with the CHS languages. The tense/aspect criterion, by contrast, would typically admit forms such as optatives and subjunctives (tenseless in many languages) as "nonfinites", forcing one to raise such counterintuitive questions as whether a subjunctive should count as infinitive or as verbal noun. The subject-coding criterion calls such forms finite and thus removes them from consideration, appropriately from an IE/HS perspective. Only for feature [15] do the IE and HS languages fail to provide a useful prototype: the finite use of nonfinite forms is predominantly the province of clause chaining, which is not a characteristic feature of either IE or HS. On the subject-coding approach, clause-chaining forms will be classified as either finite or nonfinite, depending on whether they code subject or not. This may seem an unnatural categorial split, the more so since many languages have clause-chaining forms of both types; but many languages also show greater or lesser coding of tense on different clause-chaining forms in the language, thus potentially setting up an equally artificial split under the tense/aspect criterion. Verb forms presented as "nominalized" in grammar books will normally be judged as finite if they regularly preserve subject coding; in such cases the "nominalization" will be judged to apply to the clause as a whole, rather than to the verb per se.

There is a second and more practical reason for giving preferential status to the absence of subject coding: it is a simpler criterion. In operational terms, it is a relatively straightforward matter to judge the presence or absence of subject coding on the basis of descriptions in grammar books, and to do so objectively and reliably. Tense/aspect, by contrast, is one of the most notoriously difficult and fluid subdomains in linguistics, and even specialists routinely disagree on the "right" way to envision tense/aspect in their particular language or family. Adopting it as criterial for nonfiniteness would demand a more or less detailed look at the tense/aspect system of every language in the database. And since "nonfinite" forms sometimes do show some degree of residual or reduced tense/aspect marking, this approach would also require subtle classificatory decisions about what should count as "normal" vs. "reduced" (or nonexistent) tense/aspect coding in concrete cases. None of this would otherwise be necessary, for tense/aspect per se plays a negligible role in the list of CHS resemblances⁵ and could otherwise be totally ignored in this study --- as will in fact be done. Only in a language where verbs never code subject will tense/aspect be appealed to, as a secondary criterion. See discussion of feature [10] in Chapter 5.

In other cases of reanalysis, I sometimes had recourse on a language-specific basis to the notion of a "hedged category" --- specifically, hedged articles and hedged adpositions. A hedged category

⁵ Feature [12] (Prepositional Periphrastic) does make reference to progressive aspect, but is primarily specified as a structural configuration.

resembles the corresponding true category without matching it in every particular; it provides a convenient classificatory pigeon-hole for elements which are article-like without being real articles (viz., previous-mention demonstratives) or adposition-like without being real adpositions (viz., relational nouns or adposition-like case endings). In such cases, to insist on a definitive yes/no decision on category membership would distort the essentially intermediate nature of the phenomenon: why should categorially intermediate behavior be forced into one box or the other? The hedged nature of the category, moreover, will be reflected in hedged scoring (see Chapter 5): the normal scoring range of +1 to -1 will be compressed to a range of +1/2 to -1/2 if a category involved in the definition of some feature is realized as a hedged category. Hedged categories will be discussed individually in the following paragraphs, and where appropriate in Chapter 5.

Definite articles (feature [8]) posed numerous problems of categorial interpretation. The notion "definiteness" has been explored to a considerable degree, and useful definitions are available (e.g. Chafe 1976:38ff., in terms of identifiability). But in the descriptive grammatical traditions of various languages and language groups, a wide range of adnominal particles have been enshrined as "articles". These show varying degrees of semantic overlap with prototypical "definiteness", sometimes more, sometimes less. In almost all cases the grammar book did devote some discussion to the semantics of the article, but this did not solve the basic problem: where was I to draw the line as to what was and was not to count as an "article"? And conversely: when a demonstrative particle, presented as part of the demonstrative system and not as an article, had the semantics "previous mention", was it

right to count it as an article? Such questions would be difficult enough even for someone with expert knowledge of the given language. Nor was it clear that "definiteness" per se ought to be the critical factor here; what seemed more important was the article's status as a high-frequency particle-adjunct to the noun. (The issue, of course, was where such a particle should be positioned in genitive embeddings.)

In fact I decided, in almost every case, simply to accept the grammar book's identification of an adnominal particle as "article", even when there was an explicit statement that definiteness had little to do with the particle's semantics (e.g. Fijian). Only in cases where an "article" was clearly presented as an indefinite article did I (somewhat arbitrarily) ignore it; often such "articles" were hard to distinguish from the numeral "one", and grammar books tended to describe them less fully than definite articles. On the other hand, there were a number of languages for which I reanalyzed an article-like demonstrative as a "hedged article"; see discussion of feature [8] in Chapter 5.

The notion "clitic" was another concept for which I relied heavily on the characterization presented in the grammar books. Within the present study, the relevant issue was whether an argument of a verb or adposition should count as being marked on its Head (polypersonal verb, conjugated preposition) when that argument was not affixal but a clitic. What was meant by "clitic", of course, differed from language to language, and few grammars presented explicit criteria for clitichood. Fortunately, the disparate applications of the term usually involved a rough common denominator: in almost every case it was clear, whether through explicit discussion or through examples, that the clitic did

involve some degree of phonological combination with its host word, either segmental fusion/reduction or treatment as a single accentual unit. In those rare instances of clitics where this was not readily apparent (e.g. Mixtec), I followed the usage in the grammar book uncritically. Only where the notion "clitic" was explicitly presented in purely positional terms⁶ did I not accept the alleged "clitic" as such.

Finally, there were categorial problems with adpositions. For one thing, some languages (notably Mayan) make heavy or exclusive use of relational nouns to fulfill prototypical adpositional functions --- that is, noun-like elements whose syntax is identical to that of Noun-Genitive embeddings ("table's topness" instead of "on the table"). These will be counted herein as "hedged adpositions". More subtle is the potential confusion, both terminological and conceptual, between adpositions and case endings. Such confusion typically arises in a language (Basque, Quechua, Sumerian, Yagua) having a relatively large inventory of elements described as enclitic postpositions, some fulfilling grammatical functions and some the more usual adpositional functions (local, temporal, causal). Such elements do not assign case to the nouns they attach to, but themselves constitute "case markers". Structurally, the ensemble of such elements clearly comprises a single category, for which two extreme analyses suggest themselves: either the entire set is taken as a set of case endings, or the language is

⁶ Thus for Tagalog: "An enclitic is a word that obligatorily occurs in a position after some other word or group of words" (Schachter and Otanes 1972:183). See further the Tagalog write-up in Appendix 2.

described as caseless and the endings taken as a set of postpositions. In the absence of further data, either approach may be defensible;⁷ but the choice itself appears inherently artifactual. The phenomenon is simply intermediate. Again I will sidestep the problem by resorting to a classification as "hedged adposition". By stipulation, hedged adpositions (of both types) can at best yield scores of $\pm 1/2$ for the features in which they figure --- viz., feature [1] (Conjugated adposition), feature [4bc] (Adpositional RC1), feature [12] (Adpositional periphrastic), and in a minor way feature [2b] (NP-level word order).

4.4 The language sample

The sample of 70 languages used in the present investigation comprises 58 non-CHS and 12 CHS languages: the language names are listed in Appendix 1 in two formats, first in alphabetical order (with sources), then sorted out geographically and genetically, and finally displayed graphically (Map). The 12 CHS languages are not all full-fledged members of the database. Four of them (Modern Cairene Arabic, Amharic, Coptic, Modern Irish) are later forms of older languages (Classical Arabic, Geez, Egyptian, Old Irish),⁸ included largely for diachronic perspective; these modern forms will be considered "supplementary languages", and will be ignored except for diachronic purposes

⁷ A more fine-tuned analysis might have yielded a principled way of choosing between these alternatives on a language-by-language basis. But this would have been a research project in itself, and one for which grammar books did not always provide sufficient relevant information.

⁸ Strictly speaking, Amharic is more probably a "nephew" than a lineal descendant of Geez (Classical Ethiopic) (Cohen 1936:13).

(sec. 6.5). The remaining 8 include 4 Semitic languages, which heavily overrepresents a single small family. Accordingly, in global counts and analyses only the most and the least "CHS-like" of the Semitic languages --- namely, Arabic and Akkadian --- were retained in the basic sample. Hebrew and Geez were again relegated to "supplementary" languages, bringing the total of supplementary languages to 6. The two Semitic languages retained (Arabic and Akkadian) balance the two Celtic languages, Irish and Welsh; these, with Egyptian and Berber, yield a total of 6 CHS languages included in the sample proper. With the 58 non-CHS languages, this brings the size of the core sample to 64. Schematically:

Non-CHS sample: 58 languages

Core sample: non-CHS + Irish, Welsh, Egyptian, Berber, Arabic, Akkadian

Supplementary sample: core + Hebrew, Geez, and four "recent" languages:

Modern Irish, Coptic, Modern Cairene Arabic, Amharic

The core sample is also broken down geographically into 8 areas, as follows (count indicates number of languages in the area): Europe 8, Near East 9, Africa 14, Eastern Eurasia 7, New Guinea & Australia 7, Oceania 3, North America 10, South & Central America 6. The areas are abbreviated, respectively, as EU, NE, AF, EA, NG, OC, NA, SA; these abbreviations accompany the language names on the master data table (Table 1). The Near East was singled out as an area chiefly because two non-CHS languages of the Near East (Abkhaz, Persian) showed relatively high scores for degree of CHS-ness. It therefore seemed appropriate to investigate whether the Near East as a whole, like Africa, was a high-scoring region (see sec. 6.4). Additionally, the Near East serves as a

handy buffer zone separating Africa from Europe and Eastern Eurasia.⁹

--- In the section on inter-feature correlations (6.6.1), it proved useful to lump these areas into 4 global zones containing approximately equal numbers of languages, but excluding the CHS languages. This yielded 5 groups, impressionistically rather distinct (except for EA/NG/OC) in areal/typological profile:

CHS (6) EU/NE (13) AF (12) EA/NG/OC (17) NA/SA (16).

The sample was a convenience sample (in the sense of Bell 1978),¹⁰ with certain biases emerging from the nature of the study itself (see below). As with any study purporting to make generalizations about favored and disfavored patternings across the world's languages, geographical and genetic diversity was of course a prime desideratum. In terms of simple geographic spread, the sample is fairly evenly distributed over the world (see Map). There are several exceptions. South America is heavily underrepresented, as is north-central Asia. This is due primarily to the fact that I do not know Spanish or Russian,¹¹ secondarily to the fact that relatively few South American languages have been described at the level of detail required for this study.

⁹ Note that Modern Cairene Arabic is assigned to the Near East, like its parent Classical Arabic, whereas the geographically coterminous Egyptian and Coptic are assigned to Africa.

¹⁰ Many of the phenomena examined here are "exotics" --- unusual phenomena which appear not to occur in many languages. In his article on sampling, Bell explicitly ignores investigations involving rare features of this sort: "Certain interesting special topics, such as sampling for rare types and testing for the existence of differences between languages, language groups or language types, are not treated at all" (1978:125).

¹¹ For one language, the Paleosiberian isolate Gilyak, English grammatical sketches were inadequate. Here I had no choice but to use Russian sources, which would have been impossible without several days of detailed assistance from David Peterson. I am grateful for the help.

Several languages are bunched rather closely in New Guinea and Meso-America, a consequence of the great genetic diversity of the former and the preponderance of verb-first families in the latter (see below).

Diachronically, almost all the non-CHS languages in the sample are modern languages; the only exceptions are Classical Greek, Hittite, and Sumerian. On the other hand, all the CHS languages in the core sample are old, with the unavoidable exception of Berber. Of the four "later-stage" CHS languages, three are modern; Coptic is ancient, but much more recent than its parent language, Middle Egyptian.

In terms of global genetic diversity, the sample is uneven. With only sixty-odd languages, of course, not all low-level family groups (on the order of Semitic) could be included. The genetic coverage of North America and Eurasia seems impressionistically to be more or less appropriate; note that the geographic blank spot in north-central Asia does not involve a corresponding genetic blank spot, as the major language groupings of the former USSR are all represented. South America, New Guinea, and northern Australia (non-Pama-Nyungan) are heavily underrepresented; Africa and the Near East, by contrast, have been more densely sampled, for reasons which will be discussed below.

In terms of local genetic redundancy, I have deliberately included the two major Insular Celtic languages and the two most extreme Semitic languages (as discussed above). These groups have a special status in the investigation, which made me reluctant to suppress too much information concerning them. Irish and Welsh are far from structural clones with respect to the CHS phenomena, and the Semitic languages too show considerable diversity; there seemed no fair basis for choosing Irish

over Welsh (or vice versa), or for singling out any one Semitic language as representative of the family. The procedure adopted is a compromise, minimally overrepresenting Celtic and Semitic in the sample while doing justice to the full scope of the data. --- Except for these two groups, the sample does not take more than one language from the same low-level family. Outside of Africa and Europe, it strives to avoid taking two languages from higher-level genetic groupings. The core sample has 8 Indo-European languages (including 2 Celtic), 6 Niger-Congo (including one Bantu and one Grassfields Bantu), 6 Afroasiatic (including 2 Semitic), and 3 Eastern Sudanic. Elsewhere, the closest genetic couplings are:

Hawaiian, Fijian, Tagalog (Austronesian)

Chinese, Lahu (Sino-Tibetan)

Shoshone, Pipil (Uto-Aztecan)

The selection of languages was sometimes deliberately nonrandom. A number of languages were chosen because of some particular structural feature or some link with the history of research into the Celtic/Hamito-Semitic problem, notably when previous scholars had suggested that a particular language, area, or type might be implicated in the problem. Thus a disproportionately large number of verb-first languages were included (12 non-CHS languages), including two languages relatively closely related to one another (Hawaiian and Fijian), to check the hypothesis that the CHS type might correlate with verb-firstness. Gilyak was included because it is a well-known example of a language having initial mutations. French was included specifically as an IE language having a recently evolved polypersonal verb; Classical

Greek, as an old IE language having a definite article. An areal link of the CHS type with various African languages has often been suggested, notably Bantu and West African languages; relevant here is the old suggestion that Hottentot has some especially close affinity to Hamito-Semitic (recall sec. 2.3.1.4). Hence I strove for dense genetic coverage of Africa relative to that of (say) the Americas. For similar reasons the Near East is well-represented. Basque is regularly mentioned in discussions of the CHS problem, French sometimes, Caucasian languages not infrequently;¹² they all appear in the sample.

As mentioned above, the quality of available grammar books was a crucial factor in choice of languages. I particularly sought out grammars from the Croom-Helm series or its predecessor, the *Lingua Descriptive* series,¹³ whose explicit crosslinguistic/typological orientation made them ideal for the present study; twelve such grammars appear in the sample (with Cairene Arabic making a thirteenth). In one case this involved using a dialect which is atypical of the family as a whole: Imbabura Quechua (described in a *Lingua Descriptive* grammar) is much less headmarking than are most Quechua dialects (Cole 1982:5ff.). In a different vein, Lahu was selected as a representative language from Southeast Asia owing to the excellence of its grammatical description, even though its distant relative Chinese is also in the sample.

¹² Caucasian languages are mentioned in, e.g., Pokorny 1949:242 (Abkhaz), Wagner 1959:48-54 (Avar, Georgian); French, in Wagner 1959:153-56, 173-75.

¹³ See Comrie and Smith 1977 for detailed presentation of the format such grammars are constrained to follow.

In a similar way, certain languages were singled out for inclusion at least partly because I had a linguist friend who knew the language well --- notably Hungarian and Wolof (rather than Fula). Though I almost always based my characterizations of the languages on published grammars,¹⁴ I often supplemented this information by consulting experts; their names appear in the headers to the language summaries in the Appendix. The Algonquian language in the sample (Cree) could not have been included without this expert information.

One final remark must be made, on a point of detail. It would sometimes happen that a grammar book presented data on several dialects of the given language (notably Mixtec, Basque), or on several "lects" of a language (written vs. spoken Tamil). The focus of such descriptions, however, was always some one particular dialect, and it was this that I restricted myself to in assessing the language and assigning scores. Relevant information about other dialects may sometimes be found in the mini-sketches in Appendix 2.

¹⁴ Exceptions are English and spoken French, where I relied on my own knowledge of the languages; Ingush, where my source was a ms copy of Johanna Nichols' soon-to-be-published sketch of the language; and Ineseño Chumash and Lake Miwok, drawn from unpublished UC Berkeley doctoral dissertations.

Chapter 5: Scoring and feature design

5.1 General principles

The features¹ laid out in Chapter 3 represent the point of departure for the typological survey carried out in the second half of this dissertation. Within the narrow focus of Hamito-Semitic, Celtic, and Indo-European, the phenomena seem clear enough, and it might seem a fairly straightforward task to expand the field of view to take in languages all over the world and ask whether a given feature is or is not present in each. In practice, the project is anything but straightforward. A host of difficulties arise --- some specific to a given feature, some global; some amenable to more or less objective resolution, some inherently subjective. The present chapter is devoted to recasting the CHS features from a global perspective, and to setting up a system for scoring individual languages with respect to their degree of resemblance to CHS. I will begin, in this introductory section (5.1), with general matters; feature-specific issues will follow (sec. 5.2).

¹ Henceforth the term "feature" will show a systematic ambiguity. With reference to the CHS languages it will continue to refer to the characteristic CHS phenomena, just as in Chapter 3. In global typological perspective, however, it will refer to a higher-order "feature" encompassing the relevant CHS phenomenon as one possible realization. These terms and concepts will be discussed below. The second usage will strongly dominate in the coming chapters.

I have not attempted in this chapter to give criteria covering every contingency which could and did arise for all the languages, but rather overall guidelines of general applicability. To the degree possible, the scoring reflects a painstaking attempt to set up clearcut criteria and apply them consistently across languages.

5.1.1 Privative vs. equipollent features

There are two conceptual steps in carrying out a typological investigation such as the present one, whose takeoff point is a "local" phenomenon (or phenomena) which --- in advance of the investigation itself --- has been noted anecdotally in a small cluster of languages. First, the phenomenon must be set into opposition with something else --- that is, it must be recast as one alternative realization of some higher-level typological feature. This higher feature, conceived as a "universe dimension" of crosslinguistic variation, will be said to have various feature values, one of which is precisely the phenomenon which triggered the investigation (here, the CHS phenomenon). Secondly, to each of these values will be assigned a numerical score, indicating how well that value can be taken as agreeing with the value realized in the CHS type. The central concern of this chapter is the second step, the scoring. But something must first be said about the features themselves. For specifying the set of features to be used is not something automatic or trivial. The CHS phenomena themselves are all that is given in advance; the features are not.

Linguistic oppositions are standardly considered to fall into two types, termed privative and equipollent. A privative opposition

definitionally involves the presence vs. absence of a given phenomenon, and its values are "yes" (marked) and "no" (unmarked);² an example is the feature "having clicks". Typically the marked value will be the rarer of the two. Only the marked, positive pole of a privative opposition has any positive reality of its own; the opposite pole represents merely its absence. By contrast, in an equipollent opposition all values of the feature have their own positive reality. The parade example is the feature "clause-level word order", whose 6 values (VSO, VOS, SVO, SOV, OVS, OSV) all have an equal claim to positive reality; it would clearly be wrong to characterize (say) "SOV" as merely the absence of SVO and other orders. --- Word order is a naturally multivalent equipollent feature. Other equipollent features may be naturally bivalent, e.g. the opposition of Verbal Noun and Infinitive. All privative features are bivalent by definition.

I have rehearsed these familiar concepts because they conceal a deep problem in feature design. Given a particular phenomenon P, there is an infinity of possible universe dimensions within which P can be embedded as a possible realization. The choice among these possible universes is not something which can be justified objectively, but must rather appeal to a "sense of the field", a presumed consensus of most of the linguistic community that this particular choice is in fact a reasonable one to make. In particular, of all possible design choices of higher feature, there is one logically simplest option which is always available: the phenomenon P can always be cast as the marked pole

² In some approaches the unmarked member of a privative opposition is presented as being neutral with regard to the presence/absence of the given feature.

("yes") of a privative opposition. The question immediately arises: when is this default option appropriate --- and why?

There are various reasons for having recourse to a privative feature. For one thing, a privative approach may reflect nothing more than a lack of familiarity with, or insight into, the higher-level domain D implied by a phenomenon P. We may not understand this domain well enough to set up an insightful equipollent opposition within it. If the domain were better understood, the simple privative value "no" might be split into several distinct positively defined phenomena $P'(i)$, collectively making up (together with the original P) an equipollent set of realizations of D.

Even if an obvious equipollent universe dimension does suggest itself, one may prefer a privative feature for other reasons. For example, the individual phenomenon "having a birthday on October 18" has a natural equipollent universe dimension: the set of all 366 possible birthdays. But for certain birthdays, less "anonymous" than October 18, it may be insightful to oppose that particular birthday to the aggregate of all others combined, thus yielding a privative feature. Scholars might wish, for example, to study the possible psychological effects on children of having an unusual birthday like Christmas or February 29; here a privative division of the population into individuals who are born on Christmas (or February 29), and those who are not, is clear and reasonable. In a linguistic vein, the phenomenon "having clicks" (adduced above as a privative feature) could easily be recast as equipollent: the universe dimension might then be taken as something like "all sets of consonants sharing a + value for some distinctive feature",

with clicks now being just one ordinary subclass among many. Linguists are quite aware of this, but nonetheless in this case a privative opposition seems compelling. Clicks are clearly special, and merit being profiled as a phenomenon in their own right. In the same way, one might investigate Object-first languages (OSV, OVS) as one pole of a privative feature, set off against all other word orders lumped together. This too is reasonable; again, the starting phenomenon is felt as something special. This feeling of specialness, which may be purely intuitive (based on experience and judgment) or may be theoretically informed, is what is centrally at issue in these examples, for it constitutes the most fundamental justification for setting up a privative opposition.

Yet another factor may come into play in the selection of an appropriate universe dimension. The phenomenon P may have multiple aspects to it, any of which can implicitly define a different conceptual domain D within which P can be embedded. The prepositional periphrastic construction [12] is an instructive example. One can view the construction formally, in which case a privative feature seems appropriate: does the language have this structure or not? But one might also take a functional view. Usually the Prep Periphrastic expresses progressive aspect; hence one might suggest an equipollent feature, "formal techniques for rendering progressive aspect", with the Prep Periphrastic now appearing as one such technique among many.³ The same might perhaps be said of the "DO" periphrastic, which could be taken as one of several ways of conveying emphatic semantics. Obviously neither of these approaches is in principle "right" or "wrong". They do, however, pick

³ This is in fact the approach of Blansitt 1975.

out different sets of languages, and as such may be more or less useful to a particular program of research. In the present case the CHS languages themselves indicate the more fruitful approach, and the one which was actually adopted --- appropriately, as these languages represent the take-off point for the investigation and rightly constrain the parameters of the analysis. In fact the CHS Prep Periphrastic need not convey progressive semantics (though it very often does); indeed, Egyptian and Celtic offer a choice of prepositions, with the choice conveying variable information about tense. Similarly, the Welsh "DO" periphrastic is not emphatic in the sense that English "I do love you" is. It is just a lucky accident, however, that in these cases the facts happen to favor one or the other analysis. The general "multiple-aspect" problem remains.

The above discussion may convey the impression of hairsplitting for its own sake. Yet the issue is of fundamental importance for the present study. As remarked in Chapter 1, the question of how skewed a feature's distribution is --- how strongly a minority value can be seen (or not) as bucking a clear global trend to the contrary --- is vital to evaluating how likely the given feature-value is to have arisen simply by coincidence. A strongly "skewed exclusive" distribution is maximally unlikely to represent coincidence. The skewing, however, depends on the feature, and in particular on the kind of feature. Crucially, bivalent features (and privative features, inter alia) have the maximum potential for skewing. Conversion of the feature to multivalence, implying a more fine-tuned analysis and hence a multiplicity of new subtypes, will nor-

mally lessen the skewing,⁴ splitting the majority spike into several smaller subpopulations. Hence the decision to use a privative feature (automatically bivalent) builds into the feature an automatic propensity for maximal skewing. A schematic example will make this clear. Suppose that we have a sample of 100 languages, and that a phenomenon P1 can be embedded within a higher-level feature either privatively or equipollently, the latter breaking the value "no" into three different subphenomena (the count in parentheses indicates the number of languages having the given feature value):

Privative: P1 = yes (20) P2 = no (80)

Equipollent: P1 (20) P2' (30) P3 (25) P4 (25)

In the first case, the skewing ratio (highest over lowest) is 80/20, or 4 to 1; in the second, 30/20, or only 1.5 to 1. The first represents moderate skewing; the second, almost none. We will return to this issue in sec. 5.1.2.

We will see in Chapter 6 that roughly half the features in the sample, all of them privative, show strong skewing and are properly to be counted as "exotics". In all these cases the dominant majority feature-value, and the source of the strong skewing, is the unitary feature-value "no". Accepting these highly skewed figures requires accepting the initial decision to set the features up as privative. And this in turn will rest largely on an intuition that each of the given phenomena is special enough to be worth looking at in its own right, rather than as "just another realization" of some equipollent feature.

⁴ I say "normally" because the minority value, too, might be split into multiple subtypes.

In the absence of a full linguistic theory for these phenomena, this intuition can only be justified by an appeal to the collective trained judgment of the linguistic community, and by assuming (as I do here) that my judgment will more or less coincide with others'.

The features presented herein as privative are as follows:

- 4c) Presence of the "Move Adp" strategy for Prep RCls
- 5) Presence of a special RCl verb form
- 7) Presence of a suffixing/infixing alternation
- 8) Presence of the special CHS patterning of Article in Gen embeddings
- 9) Presence of nonconcord with full-NP subject
- 11) Presence of a predicative particle in "copular" sentences
- 12) Presence of an adpositional periphrastic construction
- 13) Presence of a "DO" periphrastic construction
- 14) Presence of Adv clauses formed by "and" + finite clause
- 15) Presence of nonfinite forms used for main-clause predication
- 16) Presence of syntactically controlled word-initial change
- 17) Presence of the peculiar CHS lexical structure "kin of Gen"

For all of these the unmarked value ("no") is the commoner, usually by a large margin. Note that features defined in terms of a numerical cut-off point on a scale can always be presented as either privative or equipollent. Two such features appear in this study, viz. [1] Conjugated Adposition (number of actants coded on the Adp: 0 vs. 1) and [6] Polypersonal Verb (number of actants on the verb: 0,1 vs. 2+). These features did not show the strong skewing characteristic of most privative features; hence I chose to treat and analyze them as equipollent.

The distinction between privative and equipollent will become crucial in attempting to determine how well two features correlate with one another crosslinguistically (see sec. 6.6). Having a large number of strongly skewed privative features causes inherent problems for such correlations. The difficulty arises because, when we look down the roster of languages, any pair of strongly skewed privative features will automatically have in common a preponderance of shared NO values, meaningless values devoid of any positive reality, expressing nothing beyond absence. No such problem exists with equipollent features, where NO has real conceptual content. We will return to this issue in Chapter 6.

5.1.2 Bivalency

In setting up the features to be used in the typological investigation below, I had to balance two somewhat contradictory goals: first, to achieve a sufficiently fine-grained descriptive grid to capture the wide variety of phenomena that actually turned up in the global survey; second, to view all of these from the standpoint of how well they resembled the CHS phenomenon, and score them accordingly. The first of these considerations encouraged the making of distinctions; the second required the suppression of distinctions. In attempting to meet these clashing goals, I set up a scoring system embodying two principles. In accordance with the second goal, the features are forced to be bivalent rather than multivalent; in accordance with the first, the scoring is a matter of degree rather than simply a yes/no decision. I will describe these principles in detail in this and the following section.

Bivalency requires each feature to be set up as having just two values. For any given feature, the paired feature values will partition languages into two contrasting subpopulations, one agreeing with the CHS type for the feature and one opposed to it, and this notion of opposite-ness will be fundamental to the scoring. A language having a feature value clearly agreeing with CHS will be scored +1 for that feature, while a language clearly belonging to the opposite type will be given the opposite score of -1. As will be evident in Chapter 6, this schema will have practical consequences in the overall scoring of a language for degree of closeness to the CHS type: all the language's individual feature scores will be totaled, with + and - scores allowed to cancel each other out.

Note that this principle of scoring has nothing to do with the actual content of features and feature values. It will apply equally to equipollent oppositions and to privative oppositions. It might seem that a privative opposition, inherently involving the presence vs. absence of a feature, might more properly be scored from 0 (absence) to +1 (presence); only the value that has positive reality, on this view, would get a non-zero score. But this will not be done here. The reason is that both kinds of oppositions have the identical effect of partitioning languages into two contrasting subpopulations, each population with its own positive reality appropriately mirrored in a non-zero score (+1, -1). For any given feature, of whatever nature, one subpopulation will have the CHS feature value and one will not, and the question at issue is whether a particular language does or does not fall into the CHS subpopulation. This question is the same regardless of whether the feature is privative or equipollent, and the scoring scheme proposed

above reflects this sameness. To put things in a slightly different way, the numerical scores can in one sense be taken as mere labels for population subgroups; on this view, corresponding subgroups ("agrees with the CHS value", "does not agree") should surely all get the selfsame label (+1, -1) regardless of what the particular feature is. These mnemonic labels are numerical, of course, but the numerical nature is of only limited significance. It will be important chiefly in arriving at overall language scores in Chapter 6.

Demanding that all features be bivalent has the indisputable advantage of making the data much more tractable, but at the cost of possible conceptual distortion. For not all features are "naturally bivalent"; at the same time, any multivalent feature F (e.g. "clause-level word order") can always be recast in bivalent terms, and in several ways. First, one can simply redefine a new bivalent feature F' as the presence/absence of any one particular feature-value V of F (e.g., F' = opposition of VSO vs. "all other word order values"); we may call this the lumping approach.⁵ Or, the multivalent feature F can be factored out into two or more distinct and intersecting bivalent features (e.g. [\pm V-first], [\pm V-final], [\pm S-before-O]) --- the splitting approach. Or, perhaps the multiple values of F can be linearized, taking the original set of discrete incommensurable values and recasting them as degrees of approximation to two extreme best and worst values, the latter now taken as defining a new bivalent F' ;⁶ on this linearization

⁵ More generally, one can partition the values of F into any two subsets, each constituting one value of a new bivalent F' . Thus if F had values F_1, F_2, F_3, F_4 , one could define a new F' with values $F'_1 = "F_1\text{-or-}F_3"$ and $F'_2 = "F_2\text{-or-}F_4"$.

⁶ This notion of quantitative degree of approximation is unavoidable

approach, the extreme values will have scores of +1 and -1, while the intermediate value(s) will be given intermediate scores (e.g. assigning SVO an intermediate score [+1/2] between VSO [+1] and SOV [-1]). It is hard to imagine how one could "objectively" decide which schema of feature analysis was best in such cases of multivalent features, or which of the three strategies for creating bivalency was most appropriate. Typological experience and intuition will be indispensable --- and subjective --- guides here.

This issue has real consequences for evaluating the significance of the data, an issue already broached in discussing privative features in sec. 5.1.1 and re-presented here from a different perspective. As already argued, it is not enough simply to ask whether a given language agrees with CHS for a given feature; just as important is the question of whether the CHS type can legitimately be seen as a minority type vis-a-vis a majority type, i.e. as "bucking a trend". But to turn a naturally multivalent F into an artificially bivalent F' by setting off one value (the CHS value) against all the others --- the lumping approach --- can lead to stacking the deck by manufacturing a trend. For example, if F has eight values, each equally common, then no single value is a minority value; but if a new bivalent F' is created by combining seven of the eight into a new macrovalue, the one surviving uncombined value (here the CHS type) will indeed be a minority bucking a trend. The trend will be purely an artifact of this lumping approach; if some other (non-CHS) value were instead to be selected as the single uncombined value, the CHS value would now belong to the majority type. anyway; see sec. 5.1.3 below.

On the other hand, if the CHS value is a minority type even in the multivalent feature F , it will remain so (if singled out as the odd-man-out) in bivalent F' , only to a greater degree. --- The second approach to bivalency, viz. splitting the multivalent F into several orthogonal features F' , is not subject to this particular form of automatic bias: if our eight-valued feature F is split into three bivalent features ($2 \times 2 \times 2 = 8$), all values of each of the resultant features F' will again be equally common. As for the third option, that of linearization of values of a multivalent feature F , it could have any effect on the minority status of a given feature value, depending on just how the several values are linearized and scored and how many languages belong to each F -type.

In light of the above considerations, each feature used in this study could potentially require individual discussion as to the naturalness of bivalency, though those features which are "naturally privative" (hence naturally bivalent) should cause little difficulty. Trouble will arise especially with regard to (1) clause-level word order, (2) polypersonal verb, (3) positioning of the article in genitive embeddings, and most acutely (4) the dense complex of phenomena entering into relativization. In fact, all three strategies for creating bivalency will be appealed to: e.g., linearization for clause-level word order, lumping for polypersonal verb and positioning of the article, and splitting to render the complexity of relativization more tractable. Moreover, as we will see in the next section, even an inherently bivalent feature can have a natural third value which must be taken into account, viz. the value "feature is irrelevant". Here linearization provides a natural solution: the value "irrelevant" is recast as intermediate

between the two "real" values, and is assigned a zero score.

There is a quite distinct problem attendant on the "splitting" approach, which will come up twice in the discussion of the individual features (word order; relativization). Splitting is an excellent way to increase the precision of crosslinguistic comparison: breaking a single feature F into more and more orthogonal components $F(i)$ makes for a better and clearer description, and one which is more easily scorable. However, it must be remembered that such a feature F (e.g. relativization) is just one of many factors entering into an overall profile of the CHS type. In the final summation process which will generate a total score measuring a language's degree of fit to the ideal CHS type, how much weight should be given to F ? A high degree of feature splitting will create more and more features $F(i)$ in place of the single feature F , each contributing its own score to the final sum and thereby cumulatively assigning to F an increasingly high weight in the overall summation. Such a feature as relativization is indeed a richly articulated area of grammar, and it seems reasonable that it should contribute several features (and several summands) to the overall profile of the language. But just how much weight is appropriate? --- There is no objective answer. It is fortunate, then, that this issue can be divorced from that of feature splitting per se: a scaling factor can always be applied to all or some of the features $F(i)$ if it is felt that F is contributing too much to the overall score. This point will be addressed in Chapter 6, when the results of the survey are examined. For now, I will proceed to split features whenever appropriate, thereby allowing a more fine-tuned crosslinguistic description and a clearer scoring system.

5.1.3 Scoring as a matter of degree

The second idea underlying the scoring system is that determining whether a language has a given feature value is frequently not a dichotomous yes/no matter, but one of degree. Thus a given value of the feature may only sometimes be present in the language; or it may be co-present with another alternative value; or a phenomenon in the language may approximately resemble the intended structure, yet with qualitative differences; or the determination may depend on how the phenomenon is analyzed; or the description in the grammar book, while pointing in a certain direction, may not be sufficiently precise to make possible a categorical yes/no judgment. Therefore the numbers +1/2, -1/2, and 0 will also be used, and scores will be assigned in keeping with the following broad guidelines, suitably interpreted for each individual feature value X:

- +1 X represents the norm, or a norm (always or usually present)
- +1/2 X is present to a lesser but considerable degree; or, a construction similar but not identical to X is present; etc.
- 1/2 X is present, but as a very minor phenomenon; or, a construction with only distant affinities to X is present; etc.
- 1 X is absent or all-but-absent
- 0 Nothing can be said one way or the other

"Half" scores in general will indicate some mitigating factor tending to moderate a full +1 score, or may represent an average of scores of competing strategies. In particular, "hedged categories" will normally earn half scores, the hedged score reflecting the hedged nature of the category.

This schema has obvious advantages: it increases flexibility, makes the scoring much more realistic, and lessens the impact of the inevitable errors in scoring and interpretation (especially reinterpretation as hedged categories), which will now most likely be off by only 1/2 instead of by 1. Conversely, the process of deciding exactly what score to assign cannot always avoid a degree of subjectivity. The distinction between +1/2 and -1/2 was sometimes particularly difficult to make. Nonetheless, I made it a policy to deliberately attempt to force the issue: I tried to assign non-zero scores wherever possible, implying an explicit decision as to whether the given feature value does (+) or does not (-) constitute a significant part of the language's grammar. Equally difficult was the decision between +1 and +1/2, notably in cases involving extremely rare features that never recurred outside of Celtic and Hamito-Semitic in exactly the form found in the CHS languages. In such cases, if a language had a phenomenon which was quite similar to the CHS one without achieving perfect identity, should it be counted as a full match (+1)? The answer varied from feature to feature, but in general I felt I should lean toward flexibility rather than the reverse. A complex phenomenon like infixing/suffixing alternation (feature [7]) or nonagreement of the verb (feature [9]) could have so many niggling conditions on satisfaction as to make a perfect match extremely unlikely; finding even an imperfect match was a rare event. In many such cases, it seemed reasonable to score a good but imperfect match as +1. The net effect is that the scores may show a tendency to over-represent marginally the languages' degree of resemblance to the CHS languages. The issue will come up repeatedly in discussion of the individual features in sec. 5.2.

The score "zero" calls for special comment. There are three distinct ways that "nothing can be said one way or the other", leading to three kinds of zero scores which will be coded in three different ways. Though conceptually disparate, all three have in common the fact that there are no grounds for assigning either a + or a - score (hence "zero").

a) For features that are scored on some (semi-)objective basis, the given procedure may sometimes yield a zero result. For example, a language's overall word order type (at the NP level) will be computed by averaging scores representing N-Gen polarity, N-Adj polarity, N-RC1 polarity, and adpositional polarity. On this basis, a language having the orders N-Adj (+1), N-RC1 (+1), Gen-N (-1), and Postpositionality (-1) will score $(+1+1-1-1)/4 = 0$. More generally, the feature may be realized via two competing strategies pointing in opposite directions, whose scores thus counterbalance and cancel each other. And occasionally I simply could not decide what score to assign. --- All zeroes of this kind will be coded as "0" (true zero).

b) There may be no pertinent data available in the grammar book(s) consulted. This circumstance in itself need not force a zero score: for some structural features it can safely be inferred that failure to discuss the phenomenon implies nonoccurrence in the language (if it did occur, the grammar would hardly fail to mention it). An example of this sort would (presumably) be syntactically controlled word-initial change. On the other hand, there will be other features that can go unmentioned in a grammar book where no such inference is legitimate, one way or the other --- such a feature as the metaphorical use of kin terms, which a

grammarian might not think to discuss in a grammar book even if it were present in the language. Or the grammar book may state that a certain phenomenon occurs, yet provide no concrete information showing how it occurs. All such zeroes will be coded as "N" (for "No data").

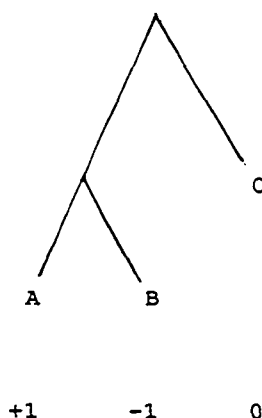
c) The feature may be conceptually irrelevant, because the language lacks a category presupposed in the very statement of the feature. Some examples are: conjugated adpositions in a language lacking adpositions; VO vs. OV word order in a free word order language; positioning of the article in genitive embeddings ("house the-boy") in a language having no article; Verbal Noun vs. Infinitive in a language lacking verbal-abstract nominals altogether. Illustrating with the feature VN/Inf, we thus get three ideal types:

Type A: Verbal Noun (the CHS value)	+1
Type B: Infinitive (the opposite value)	-1
Type C: No verbal abstracts (feature irrelevant)	0

Such zeroes will be coded "I" (for "Irrelevant"), and represented on the data tables with a blank (" ").

Type C is a problem, and it is not self-evident what is the best way to treat it. As remarked above (sec. 5.1.2), the value "irrelevant" can be seen as a third feature value, automatically converting a bivalent feature to a multivalent one; what then should be done to restore bivalency? Assigning Type C a zero score, as we have done, amounts to invoking a linearization approach to the problem. One could also consider a lumping approach, with Types B and C grouped together: after all, both represent ways of failing to realize the CHS value (Type

A). In principle a splitting approach might seem best of all, for logically Type C does not stand in opposition to either A or B separately, but in meta-opposition to the two together:



Properly speaking there are two different features or dimensions at issue here, the first a "metafeature" vis-a-vis the second, which is logically subordinate to it:

(i) Presence of verbal abstracts (yes, no)

(ii) VN vs. Inf (for languages having the value "yes" for [i]).

In a hypothetical "total typological statement" of all the structural possibilities inherent in the world's languages, these two dimensions would indeed be split apart, and each would merit separate treatment. But this is an overrefinement in the context of the present study. We are concerned here only with features which represent a structural wedge dividing CHS and the bulk of Indo-European, and only feature (ii) --- the opposition of A vs. B --- is relevant in this regard. In fact CHS and Indo-European almost always show agreement regarding the "metafeature" (i), for almost all these languages do have verbal abstracts. The splitting approach, then, will not be adopted here; the

problem thus becomes how best to incorporate Type C into a scoring system designed around a single feature.

As remarked, one could lump Type C with Type B into a single macro-value, encompassing those languages failing to realize the CHS value (Type A). There is a certain intuitive attractiveness to this; indeed, in a sense type C could be taken as conceptually even further away from Type A than is Type B. Note, however, that for many features the subpopulation of languages represented by Type C is rather large; as argued in sec. 5.1.2, the effect of this lumping could thus be to artificially enhance the degree to which the CHS type (Type A) appears as a minority type. And of course such a lumping procedure is arbitrary, for Type C could with equal justice be lumped with Type A. --- The linearization approach, in this case, seems more natural, especially since the appropriate intermediate score is obvious: zero. This score, neutral between +1 and -1, mirrors the conceptual neutrality of the notion "irrelevant". In the final summed assessment of a language's overall closeness to the CHS type, a score of zero will contribute nothing one way or the other. Type A and Type B will be equally free from artificial inflation. We will return to this issue in Chapter 6.

5.2 Features

1) Conjugated adpositions

a) If the language lacks the category adposition (as often in Australia, North America), score zero ("I"). If there is only a tiny handful of "adpositions" (say 1 to 3), the scoring will depend on whether

these adpositions figure centrally in the language's grammar: if yes (Fijian), the category "Adp" will be taken as present; if not (Hua, Cree), as irrelevant, with a zero score ("I"). Compound adpositions ("on top (of)") will be ignored.

b) To be scored +1, the combination [Adp + PronObj] should count in some sense as comprising a single word. This can involve either affixal object morphs or clitic object morphs, depending on the language; see sec. 4.3 for further discussion, and cf. feature [6] below (Polypersonal verb). If adpositions can take pronominal objects either as clitics or as free forms (Mixtec), score +1/2.

c) An adposition can count as "conjugated" even when the form of the PronObj is segmentally identical to the free pronominal form, if the PronObj acts like a clitic.

d) Conjugated adpositions, of course, involve pronominal objects; but in some languages a pronominal copy regularly appears on the adposition even when the object is a full NP ("with-him John"). This difference will be ignored here (though it will play a role in the treatment of Relative Clauses, [4] below). Cases of this sort will count as ordinary conjugated adpositions.

e) Some languages have relational nouns, which function like adpositions but whose syntax is identical to that of N-Gen embeddings ("table's topness" instead of "on the table"). Relational nouns will be counted and scored as "hedged adpositions" (see [g] below). Languages so analyzed include Eskimo, Koasati, Mixtec, Pipil, and Tzutujil. Some of these languages lack true adpositions entirely (Eskimo, Koasati); in some, a few true adpositions exist (see [g]). On the other hand, not all languages lacking adpositions need have a clearly profiled notion of

"relational noun". There was undoubtedly a degree of crosslinguistic inconsistency in my assignment of the descriptive label "relational noun" to adposition-like elements in particular languages, partly reflecting the descriptive tradition embodied in the grammar book; note that the reanalyzed languages listed above all come from the New World.⁷ In general, a language would not be reanalyzed if the number of such quasi-adpositions listed in the grammar was small, or the number of true adpositions large. Any errors, fortunately, should not change the scoring by more than 1/2 for a given feature.

f) There is a conceptual difficulty when the adposition itself is presented as a clitic (typically an enclitic postposition). As remarked in sec. 4.3, these too will be analyzed as "hedged adpositions"; languages so reanalyzed include Basque, Quechua, and Sumerian.⁸ It would seem that such Adps must ipso facto count as "conjugated", because the combination [PronObj + Adp] will automatically involve a clitic bond. However, so too will the combination [NObj + Adp], and in exactly the same way. One could, technically, count such cases as "conjugated Adps"; but to do so merely on this basis would be out of keeping with the specifically pronominal nature of the prototypical phenomenon (as found in the CHS languages, the takeoff point for this investigation). I will not do this. Rather, I will count the combination as

⁷ Several West African languages, such as Gbeya and Yoruba, might perhaps have been so analyzed but were not. In West Africa and Southeast Asia, the picture is further complicated by deverbal adpositions.

⁸ Postpositions in Yagua, apparently constituting a single form-class, display mixed phonological behavior: some show no clitic attachment to full-NP Objects, while others are always clitic-like and hence count as "hedged postpositions". The language will be analyzed here as having full-fledged postpositions.

"conjugated" only if there is some clear evidence of phonological fusion or reduction over and above the general clitic nature of the Adposition.

g) Scoring for hedged adpositions (of either kind) cannot go outside the range of $\pm 1/2$. If the language has both true and hedged adpositions, the score assigned (not only here but for other features defined in terms of adpositions) will be an appropriate compromise reflecting the relative importance of the two types.

h) Sometimes only certain adpositions show conjugatedness; here score $+1/2$ if conjugatedness seems a major strategy in the language, $-1/2$ if not.

i) If only certain person/number combinations involve conjugatedness, score as $+1/2$ if more than half of all possible combinations are conjugated, $-1/2$ if less than half, 0 if roughly equal.

j) The "of" particle of Bantu will not be counted as a preposition.

2) Word order

a) Word order will be split into two subfeatures:

(2a) Clause-level word order

(2b) NP-level word order⁹

There are several reasons for doing this, despite the known correlations

⁹ Matthew Dryer has shown that the widely assumed correlation between Adj-N order and OV order does not hold (see e.g. 1989:274). This might argue for splitting off Adj-N order as a third, separate word-order subfeature alongside (2a) and (2b) above. I will not do this, however. To factor word order into three features would overrepresent the phenomenon when set alongside the many other features under investigation, and in the present context would amount to needless precision. Further, the category "Adjective" is problematical in many languages, and profiling it in a separate feature would require detailed decisions about scoring, "hedged" adjectivity, etc. --- none of which would otherwise be neces-

between the two:

i) Like relativization, word order properly counts as a macro-domain, covering considerably more conceptual "territory" than the other features surveyed. Since relativization must unavoidably be split into several subfeatures, it is not unreasonable for word order to be as well.

ii) Languages often differ in word order behavior at the NP level and at the clause level, notably as regards rigidity of ordering. Even in a language with fairly strict clause-level word order, some NP-level parameters may allow flexibility (Tagalog); conversely, a language's word order may be free at the clause level, but constrained at the NP level (Hungarian, Yimas). Separate scoring allows such differences to emerge clearly. In particular it permits either of the two levels to be given the value "irrelevant" (zero) independently of the other.

iii) NP-level word order features are naturally bivalent; clause-level word order is not (6 natural possibilities). Hence the two cannot be treated analytically on a par.

iv) In particular, the difference between VSO and SVO word order, a potentially significant matter, is purely a clause-level phenomenon. This difference is unlikely to emerge clearly from the

sary. Finally, it would be pointless to strive here for completely independent features; the features used must be those which descriptively capture the CHS phenomena, and some of these will inevitably show correlations. Lumping Adj-N order in with the other NP-level factors is convenient, and little is lost; the worst possible "error" will be $\pm 1/2$. But error is not really what is at issue, just a blurring of information; for one could, if one wished, describe feature (2b) as simply a weighted average of an AdjN score and a conglomerate GenN/RC1N/Adp score.

scoring scheme unless clause-level word order is split off as a distinct subfeature.

v) Finally, there is a straightforward way to assign a score for NP-level word order, by assigning + or - scores to the four parameters N-Adj, N-Gen, N-RC1, and Adposition, then averaging; working with four items yields scores conveniently spaced at intervals of 1/2. If clause-level word order were to be superadded to such an averaging schema, it is unclear how much weight it should be given vis-a-vis the four NP-level parameters.

b) At the clause level, code as follows:

VSO,VOS (V-first)	+1	SVO	+1/2
SOV,OSV (V-last)	-1	OVS	-1/2

As is known, SVO languages tend crosslinguistically to pattern rather like VSO, and the scoring reflects this: all VO languages will have a + score, but with SVO as a weaker variety. Perhaps arbitrarily, the same procedure is applied to OV languages: all OV languages are have a - score, with OVS as -1/2 (only Hixkaryana). Thus the + or - polarity mirrors the ordering of V and O. Note that this procedure has the effect of linearizing a multivalent feature and recasting it as a bivalent feature, "V-first vs. V-last"; V-medial is assigned an intermediate (though nonzero) score.

c) If the language is described as having basically V-final order but with considerable flexibility regarding the possibility of postverbal argument placement, code as -1/2. (If the flexibility involves only the scrambling of preverbal NP order, assign a full -1 score.) In

principle, the same ought to apply to "flexible VSO" order, but here the normal alternate order, SVO, is itself already scored +1/2; the scoring system is not fine-grained enough to code the relevant intermediate value. Hence dominant VSO order will always be coded +1, regardless of flexibility.

d) Some West African languages (notably Mande) have the highly characteristic word order S-O-V-Other. This type seems conceptually intermediate between OV and VO, and will be scored 0 (true zero).

e) If the language's word order (clause-level or NP-level) is characterized as basically free (Dyirbal) or pragmatically controlled (Hungarian), the relevant word-order feature is scored zero (coded "I") --- even if the grammar mentions a slight statistical preference for some particular word order. Some degree of subjectivity is unavoidable here.

f) At the NP level, the overall score is arrived at by an average of scores for the four subfeatures N-Adj, N-Gen, N-RC1, and Adpositional polarity. That is, a score of +1 is assigned if all the features show ideal VO values (CHS type); +1/2 if a majority of the features do; -1 and -1/2 for the corresponding OV values; and 0 if no clear preponderance emerges (Yagua).

g) At the NP level, if both values of a given subfeature are possible (e.g. both N-Adj and Adj-N), score the subfeature as -1/2, 0, or +1/2, depending on which possibility (if any) seems favored.

h) Regarding the order of HeadN and RC1, the internally headed and correlative types of RC1 will be scored zero ("I"), as neither has an external HeadN.

3) Relative clause linker

a) On the problem of encapsulating relative clause behavior in a few bivalent features, see discussion under feature [4] below.

b) The clause-linker feature is naturally multivalent. A RC1 linker may be zero, or an invariant relative particle, or a relative particle coding the Head Noun's case role in the matrix clause, or a relative pronoun coding the case role in the RC1 itself. It may code the gender/number of the Head Noun, while specifying nothing about case function. Or it may indeed indicate minimal case information about the Head Noun's role (in either clause), yet in a way that does not resemble normal case marking in the language (e.g., coding only a two-way distinction, "direct vs. oblique"). --- An ideal global typology would take all these types into account. Our purpose here, however, is specifically to examine the structural cleavage between the CHS type and the common Indo-European type, and this involves one clear contrast: relative pronoun (coding the role in the RC1) in Indo-European, vs. a variety of more or less invariant devices in the CHS languages. We will, then, reduce "Clause-linker" to a bivalent feature in accordance with this cleavage. Full-fledged relative pronouns will be scored -1; any particle not coding case will be scored +1; particles that convey case information but are distinct from relative pronouns will be scored +1/2 (see [c] below). This imposes a degree of linearization, while lumping together the zero, invariant, and gender/number types of relative particle. Note that the CHS type is not a minority type here, with or without lumping; zero and invariant linker-types are both extremely common. In fact it is the IE type which represents a decisive minority.

c) If the relative particle explicitly marks case, but the case role is that in the matrix clause, score +1/2 (Shoshone).¹⁰ If the language has two distinct relative linkers, with the choice itself indicating a minimum case distinction (e.g. direct vs. oblique role in the RCl), again score +1/2 (Welsh, Squamish, Nkore-Kiga). These two types resemble the IE relative pronoun in conveying case information, but also deviate from it significantly (in two different directions). The fact that the assigned scores have a positive polarity (+1/2 rather than -1/2 or 0) reflects my decision to accentuate the cleavage vis-a-vis the pure relative pronoun of Indo-European (-1).

d) If the language has no real category RCl (Hixkaryana), score zero ("I"). For internally headed and correlative type RCls, again score zero ("I"), since there is no conceptual possibility of linking HeadN and Clause (HeadN being internal to the clause).

e) Only true linker markers are at issue here, i.e., markers which come between the Head Noun and the clausal part of the RCl. Any marker of RCl-hood that occurs intraclausally, or at the opposite end of the RCl from the HeadN, will be ignored, for it does not "link" anything. (Occasionally this approach yields results that verge on counterintuitive, as in Amharic, where the RCl verb is clause-final but takes a prefixed RCl-marker yä-, clearly segmentable as an "all-purpose subordinator".)

f) See final "Caveat" to feature [5f], re analyzing a possible linker morpheme as part of a "special relative form" instead.

¹⁰ The relative particle in Classical Arabic codes matrix-clause case, but only in the dual; otherwise the particle is case-invariant. Score +1.

4) Relativization strategy

a) Relative clause formation is one of the most richly articulated phenomena in the typological repertoire, and certainly the most complex within the present study. Here there is no pretense of considering the phenomenon as a whole to be a single multivalent feature which must somehow be reduced to bivalency; obviously there is a whole complex of interacting dimensions. As elsewhere in this study, we will try to focus only on those points whereby the CHS languages differ from the bulk of IE, as laid out in Chapter 3. But reducing these "points" to bivalent features that are tractable and scorable crosslinguistically is an open-ended task. A rough threefold breakdown was attempted in Chapter 3: (3) linker type; (4) overall technique; (5) special relative form. In this chapter, too, we will consider (3) and (5) separately. But feature (4) is too amorphous to be useful without considerable refinement. More feature splitting will be necessary.¹¹

b) One of the main distinctions treated in Chapter 3 was the difference between copying and gapping relativization strategies. Here the salient difference between the CHS languages and IE emerges specifically with prepositional and genitival RCLs:

¹¹ Recall in this connection the discussion of feature splitting and feature weighting in a language's overall summed score (sec. 5.1.2).

CHS (copying)

IE (gapping)

Prep: man that [I spoke with him] man [with whom I spoke ____]

Gen: man that [I saw his wife] man [whose wife I saw ____]

By contrast, subject RCLs typically gap in both language groups; with object RCLs, gapping is standard in IE, usual in CHS:

Subj: man that [____ spoke]

Obj: man that [I saw ____] (sometimes: man that [I saw him])

Hence it seemed reasonable to focus only on strategies for genitival and prepositional (or adpositional) RCLs, ignoring Subj RCL and Obj RCL formation. A first feature split, then, will create two separate dimensions: Gen RCL technique (4a) and Adp RCL technique (4b).¹²

c) However, the apparent opposition "Copying vs. Gapping" is in fact a false dichotomy. It is not unusual, in strongly head-marking languages, for RCLs to have a pronominal copy which was already present in the underlying non-RCL:

Prep: I spoke [with-him the man] ==> man that [I spoke with-him ____]

Gen: I saw [his-wife the man] ==> man that [I saw his-wife ____]

(DObj: I saw-him the man ==> man that [I saw-him ____])

¹² Regarding Adp RCLs, one distinction which does not emerge from this scheme, even with the refinements to be discussed below, is that between gapping of Obj-of-Prep and gapping of the entire PrepPhrase:

house that [I lived in ____]

house that [I lived ____]

For this type, which I will call the Gap-Copy type, the coreferential noun is gapped and a pronominal copy exists; the pronoun is not created, but merely preserved. Thus another split is needed. I will work with two features:

(i) The negative of a feature "Does the NP gap?" The negativity is needed to conform to the general scoring convention whereby the CHS type (here the non-gapping type) must always get the +1 score.

(ii) Is there a pronominal copy present in situ in the RCl?

The stipulation "in situ" is important: it is meant to exclude cases where a pronominal copy exists but moves --- as with the Indo-European relative pronoun, or the "moving preposition" type to be considered below (f). The crucial intuition underlying the notion "Copying", surely, is whether or not the relative clause looks just like a normal finite clause, with all arguments present and appearing in the same syntagmatic position they would normally occur in. The second feature (ii) was constructed so as to reflect this conception.

The above two features yield a set of four feature assignments:

Type	- [NP Gap]	Pron Copy In Situ	
Copying	+1	+1	(CHS)
Gap-Copy	-1	+1	(Abkhaz)
Gapping	-1	-1	(Japanese, most IE)
Intact	+1	-1	(Kobon, Hittite)

Note the last feature combination (no gapping, no pronominal copy); this

implies that the full-NP argument is left intact within the RC1, something which occurs by definition with internally headed RC1s and correlative RC1s¹³ and also happens with a few languages having otherwise "normal" externally headed RC1s. --- Both these features will be examined twice over, for both Gen RC1s and Adp RC1s:

(4a)	Gen RC1	- [NP Gap]
(4A)	Gen RC1	Pronominal Copy
(4b)	Adp RC1	- [NP Gap]
(4B)	Adp RC1	Pronominal Copy
(also: (4c)	Adp RC1	Move Adp --- see [f] below)

With (4c), this yields 5 distinct subfeatures, which will be coded and entered as discrete entries in the master data table (Table 1a).

d) The above feature-pair (i, ii) lends itself naturally to combination and linearization for overall scoring purposes. Intuitively, the Gap-Copy type is conceptually intermediate between pure Gapping and pure Copying (its name was devised with this intermediacy in mind). In a different way the Intact type is also conceptually neutral, for it involves neither Gapping nor Copying. To arrive at a plausible combination score, one need only average the two feature subscores: the results are +1 (Copying), -1 (Gapping), 0 (Gap-Copy or Intact). Thus the two features can be made into one, at the cost of losing information; the score 0 will now be multiply ambiguous (covering, as always, other kinds

¹³ In Hittite, which has correlative RC1s, the intact full NP is accompanied by a "relative adjective"; but this does not constitute a pronominal copy.

of indeterminacy and categorial irrelevancy). This combination of 4a,4A and of 4b,4B yields the "compressed" format seen in Table 1b, where the features have indeed been combined; and it is these combined features which will actually be used in all subsequent analysis.

e) If a language can only relativize genitives by recasting the clause as a dative construction, or only by "Possessor raising" (Maricopa), score zero ("I") for all features involving genitive relatives. If a language can only relativize Obj-of-Adposition by building a voice-changing flag into the verb and recasting Obj-of-Adp as Obj of verb, again score zero for relevant features. Thus:

Gen RC1 by Dative:

man that [to-him I admired the car] "man whose car I admired"

Gen RC1 by Possessor Raising: (here "white" is a verb)

woman REL [[woman's hair] whites] ==>

woman REL [woman hair-whites] ==> woman REL hair-whites

"woman whose hair is white" (paraphrasable: "woman who hair-whites")

Adp RC1 by voice-changing flag:

knife that [I with-cut the meat] "knife I cut the meat with"

Assigning zero scores in this way amounts to a judgment that the language is not really relativizing on genitive or Obj-of-Adp at all. The syntactic reshufflings involved in the switch to dative or "Possessor-raised" syntax and the change in verbal voice (respectively) are not part of relativization per se, but represent processes that are

logically prior to and independent of relativization; their domain is restricted to the clause as such, and they could presumably take place (as options) even if the clause were not embedded as a relative but occurred as an independent sentence.

f) In Chapter 3 it was remarked that Berber and Old Irish prepositional RCLs have the characteristic that the bare Prep migrates to a position adjacent to the verb of the RCL.¹⁴ Within the context of generative grammar, and within a language like English whose RCL-Linker is a Relative Pronoun, similar phenomena have traditionally been dealt with in terms of Pied Piping: the RelPron is said to migrate away from its underlying position, moving to COMP and carrying the Prep with it. But in fact the phenomenon per se (movement of an Adposition) turns out empirically to occur with all sorts of linker types, even zero. It is thus not parasitic on some other element moving, but merits being treated as a feature in its own right. --- Moreover, there are several variants on the theme. Sometimes the Adp moves into the verb (incorporation) or onto it (clitic), perhaps in a slightly modified phonetic shape. Sometimes the Adp itself disappears, but is "reincarnated" as an adverbial element which appears adjacent to the RCL verb. And it may indeed happen that the Adp moves when a RelPron moves, by Pied Piping.

Thus I will assign a feature "Adp movement" (4c), with scores as follows:

¹⁴ Such movement is mentioned as an example of "headward migration" in Nichols 1986:84-85, though not in connection with relative clauses.

- +1 Adp moves to Verb and remains clearly identifiable
 as such (including incorporation or cliticization)
- +1/2 Adp vanishes, but a new Adv element appears beside Verb
- 0 ("I") Adp vanishes, but V is flagged for voice-change with a mark
 that does not resemble the Adp (see [e] above)
- 1/2 Adp moves, but only as part of more general RelPron movement
- 1 No movement: Adp either survives intact in situ or
 deletes entirely

The prototype of this feature involves movement specifically to the Verb, and movement restricted specifically to Adpositions. Here IE-style relativization presents a striking contrast: the movement is not to the verb but to the linker position ("COMP"), and the motion of the Prep is incidental to the much more general motion of the RelPron. Indeed, the only commonality is the fact that some sort of movement occurs. Hence the score -1/2 (above).

g) Because the Move-Adp type of RC1 (just discussed) does create a gap, it seems reasonable that the two Gapping/Copying features presented in (c) above should be scored just as for normal gapping.¹⁵ This has the interesting consequence that identical scores for the Gapping/Copying features will occur with the pure Gapping type (Japanese), the Move-Adp type (Berber), and the RelPron type (IE). Other features will adequately distinguish these types; in terms of

¹⁵ Conceivably, the Adp could move but the pronoun survive in situ; however, I have seen no examples of this. Note too that some languages employing Move-Adp may also have another, competing strategy for Adp RC1s, in which case the scores for features 4aA and 4bB will reflect both strategies.

gapping behavior, they are identical.

h) If the language lacks a meaningful category "Relative Clause" altogether (Hixkaryana), score zero ("I") for all features. If the language lacks the category Adposition, score zero ("I") for all features affecting Adp relatives. If there are only "hedged adpositions" (see feature [1]), scores are restricted to the range $\pm 1/2$. If the language does not allow Adp RCls or Gen RCls, score zero ("I") for the appropriate features. If the grammar makes no mention of adpositional or genitival RCls, assume they are ungrammatical (unless there are examples demonstrating the contrary); if they exist but occur only infrequently, or with restrictions, or somehow represent a straining of the language's resources, change the appropriate Gapping/Copying scores from ± 1 to $\pm 1/2$.

i) Regarding genitival RCls, several languages have the difficult RCl type

house [high (its) roof] (= roughly: "house whose roof is high"). Some languages include the possessive pronoun "its" here; others omit it. This construction can be multiply ambiguous in syntax and interpretation; is the underlying non-relative clause really

(i)	the house's roof (is) high	====>	Gen RCl
or (ii)	the house (is) high (as to) (its) roof	====>	Subj RCl
or (iii)	the house(TOPIC), (its) roof (is) high	====>	Topic RCl ?

On interpretations (ii) and (iii), there is in the underlying clause no direct genitival link (no N-Gen constituent) combining "house" and

"roof"; and if the particular language omits the word "its" (Tagalog, Japanese, Lahu), there is not even an indirect pronominal genitival link. The construction thus might or might not represent a "genitival RC1". In general, I will ignore this type completely unless it is the sole candidate for possible genitival RC1 in the language. If so, I will count it as a "hedged RC1" if "its" is present (Squamish: score between $\pm 1/2$), but not otherwise (Lahu, Tagalog: score "I").

j) If there are several relativization strategies, focus on the dominant one, if any. If all (both) strategies seem equally natural, assign a score reflecting an intelligent average.

k) Throughout, and in features [3] and [5] as well, I will generally ignore purely literary relativization strategies, especially ones borrowed from superstrate languages (as in Tamil, Basque, Turkish). Only "native" strategies, current in the spoken language, will be considered. Where the grammar book focuses on the literary standard language, however, providing little information on relativization in colloquial speech, I will follow the literary language (Georgian).

5) Special relative form

a) If there is a special verb form which is used only in relative clauses, or whose only additional use is in clefting and perhaps WH questions (taken as clefts), score +1. This applies regardless of whether the form is finite or "participial".

b) If RC1s involve the same verb forms as independent clauses, score

-1. This is intended to apply to isolating languages, too; here, of course, the same single verb form typically appears in all contexts.

c) If there is a special verb form characteristic of subordinate clauses in general (including RCls), score 0 (true zero). This type seems conceptually intermediate between (a) and (b), and its zero score reflects that intermediacy: a special form does exist, but not one specific to RCls. (In effect, a multivalent feature is here being rendered bivalent by linearization.)

d) If there are competing relativization strategies and only one involves use of a special form, code +1/2; in general, treat competing strategies by taking an intelligent average. If special forms exist only for certain tenses, certain person/number combinations, etc., score +1/2.

e) If a special RCl form exists only for a single verb (as with the copula in Japanese and Welsh), score -1/2.

f) There are cases where a clitic (or clitic-like) morpheme M, occurring medially between HeadN and RCl.Verb in the context

HeadN M RCl.Verb ... or ... RCl.Verb M HeadN ,

might be analyzed plausibly in two ways: either (i) segmented off as a Linker [3], or (ii) taken as part of the RCl verb, thus possibly creating a special relative form of the verb [5]. Schematically (for HeadN-RCl order):

(i) M as linker: HeadN M [RCl.Verb ...]

(ii) M as part of RCl.Verb: HeadN Ø [M-RCl.Verb ...]

In such cases, all other things being equal, I will favor the linker analysis (e.g. Tauya). The difficulty with analysis as a special relative form (score +1 for feature [5]) is that it automatically yields a

zero linker and thus automatically generates a second +1 score (feature [3]), thereby in effect assigning a single phenomenon double weight. Even if carried out consistently, this would amount to unjustified scoring inflation; it would also reduce the independence of the features. --- These considerations will not apply if M is presented as a verb affix (e.g. Dyirbal, Maasai), unless special considerations argue for reanalysis (as in Tauya); normally, a bound affix cannot be a linker.

6) Polypersonal verb

a) If the language codes two or more arguments on the verb, score +1; if one or zero arguments, score -1. This has the effect of linearizing the naturally multivalent¹⁶ feature "number of actants on the verb", lumping the basic discrete values (0, 1, 2, 3, etc.) into two groups. This particular bivalent cut captures a typological intuition which most linguists will hopefully share: that the yes/no question of whether a verb codes any sort of object is one worth asking, that is, that "polypersonal" is an insightful notion. Note that the danger of artificially enhancing a feature value's minority status by lumping (recall sec. 5.1.2) is lessened here by the fact that two distinct lumpings of subvalues are involved at once, {0,1} vs. {2,3+}; there is no single odd-man-out counterposed to "everything else". In point of fact, minority status is not at issue anyway: about half the languages in the

¹⁶ In the specific context of feature [6], terms like "multivalent" are polysemous. As always, the feature itself can be characterized as bivalent vs. multivalent; but one can also speak of the verb's own valency as monovalent, bivalent, trivalent, etc. In context there should be no confusion.

sample have polypersonal verbs, and most of those code two arguments.

b) Normally, the notion "actant coding on the verb" will involve fused person/number marking. In some languages, the verb splits up person, number, and possibly gender into separate morphs (with number typically optional); in such cases, focus on person coding where possible. If the verb marks only number (Shoshone), then do pay attention to number.

c) As with conjugated prepositions (feature [1]), the question of how to treat clitic arguments comes up. Of course, the clitic will have to be a plausible candidate for "actant coded on the verb": thus it should be part of the verb complex, rather than (e.g.) a second-position clitic or a movable clitic, and if it is a Subj clitic it should occur on the verb even in the presence of a full-NP subject (Fijian, Wolof) rather than being mutually exclusive with a full-NP subject (Mixtec). If these conditions are satisfied, clitics will normally be counted into the tally of "markers coded on the verb", sometimes even occurring alongside true affixal person/number markers (Albanian). --- Of course, not every pronoun that occurs (or even obligatorily occurs) with a verb need count as a clitic.

d) Almost always a monovalent verb will code only Subj; a bivalent verb, Subj and DObj; a trivalent verb, Subj, DObj, and IObj. There are a few notable exceptions. A few languages (Hottentot, Gilyak) code Obj but not Subj; Hua codes Anticipatory Subject (= subject of the next verb) as third argument. Properly speaking, "Anticipatory Subject" is an argument of a different verb, not of its morphological host verb; but such refinements will be ignored here. The issue is strictly number of pronominal arguments coded.

e) In some languages, a transitive verb may code either one or two

arguments depending on the particular choice of person/number (Maasai, Quechua), definiteness (Hungarian), tense/aspect/mood, etc. Such a language will be scored $-1/2$, 0, or $+1/2$, depending on whether mono- or bivalency of the verb predominates overall. On the other hand, if transitive verbs in the language have the choice of coding either one or two arguments "across the board", with no conditioned split evident (Persian), focus on the bivalent coding and score $+1$; similarly if some transitive verbs code one argument and some two (Amele).

f) Some languages (Gilyak, Ingush) present a theoretical potential for coding Subj and Obj together on the verb, but a potential which for various reasons is seldom actually realized: heavy lexical restrictions (Ingush); subject-coding both partial (number only) and optional (Gilyak). Score these $-1/2$.

7) Infixing/suffixing alternation

a) A score of $+1$ indicates a clear infixing/suffixing alternation of the Obj marker within the verb, with the conditioning involving a preverb or something similar (the "preverb" may be a Subj clitic, as in French). This matches the basic phenomenon seen in Berber and Irish. The alternation may apply to the conjugation of a given lexical verb, if the preverb is grammatical, or to simplex/complex pairs of verbs (one with preverb, one without) if the preverb is lexical. Considerable leeway will be given to the notion "infixing"; the only positional stipulation (in keeping with Berber and Irish) is that the two alternant forms of the Obj occur on opposite sides of the verb root. Even a prefixing/suffixing alternation (Yimas) can count as a full-fledged

example of the phenomenon. Note too that "affix" is intended here as a broad cover term, including the possibility of clitic alternation (Wolof). --- If the conditioning for the alternation does not involve preverbs but some other factor such as aspect (Sumerian) or independent vs. conjunct status (Cree), score +1/2. If the grammar makes no mention of any Obj alternation, score -1.

b) A logical prerequisite is that the verb does code Obj markers (either as affixes or clitics). This normally implies a polypersonal verb, except in the few languages where monovalent verbs code Obj as their single argument (see [6] above). If the language does not code Obj on the verb, score zero ("I"). If a language codes Obj on the verb only half the time (0 score for feature [6]), and shows no infix/suffix alternation, score -1/2 (instead of -1 or 0).

c) Several grammars present an "infixing" Obj alternation involving positional alternants on the same side of the verb. Such alternations (e.g. Amharic) readily lend themselves to reanalysis in terms of an optional position-class ("Slot") peripheral to the Obj marker:

Suffix: V-Obj "Infix": V-Obj-Slot (or mirror image).

These count only minimally as "alternations", and hence will be scored -1/2. However, two such languages (Hua, Burushaski) show a twist. Here Obj is normally prefixed, but for verbs beginning with certain phonemic sequences (ha- in Hua, d- in Burushaski), Obj is infixes after the initial element. These initial elements are essentially part of the verb stem; they are of common occurrence but have no meaning, which weakens the case for segmenting them off as a true slot. Score 0 (true zero).

d) Languages where conflated (portmanteau) Subj-Obj marking is the norm present a problem. Here the conflated nature of the marking comes

close to precluding even the possibility of alternation (movement of Obj marker would ipso facto be movement of Subj too). If such a language lacks infixing/suffixing alternation, score -1/2 (not -1).

e) Numerous languages have compound tenses involving a separate Aux word. Here the MainVerb is recast as a nonfinite form, thereby losing all or some of its argument marking to the finite Aux. Often (though not always) this process involves relocation of the Obj marker:

V-Obj vs. Aux-Obj V.

This will not be counted as an infix/suffix alternation (score -1). To be sure, the Obj marker does "change places" vis-a-vis the MainV, but the alternation occurs above the level of the individual word. Any such "alternation" involving a separate Aux word will be ignored.

f) One language (Squamish) shows an alternation of Subj (not Obj) marker conditioned by the presence of a preverb; score -1/2. Positional alternations of Subj that are not conditioned by preverbs (Semitic, Afar) will score -1.

8) Definite article in genitive embeddings ("house the-boy")

a) Much flexibility will be allowed as to what counts as an "article", with the grammar book's own usage dictating the decision almost everywhere (see sec. 4.3 for discussion of the conceptual difficulties below). Only articles clearly characterized as being indefinite will (somewhat arbitrarily) be ignored. If the language lacks articles, score zero ("I"); but see [8g] below on the possibility of reanalyzing demonstratives as articles.

b) With this feature more than most others, grammar books seldom

addressed the relevant issue (positioning of the article) at all. Usually I had to infer the behavior on the basis of examples. Where there was no discussion, or insufficient (or no) exemplification, I scored zero ("N" = No data).

c) The feature "article positioning" is not only naturally multivalent, but luxuriantly so. Straightforward article placement constraints such as Head only, Dependent only, both Head and Dept are only the beginning; one could further ask whether the article is constrained to be adjacent to its noun (cf. the "sandwich" construction" of Greek or Hawaiian, [the [the-Dept] Head]); whether it occurs to the left or the right of the noun; whether it perhaps comes at the left edge (or right edge) of the [Head-Dept] constituent as a whole, regardless of the order of Head and Dept; and one could imagine combinations of the above. Given this thicket of possibilities, it is less than obvious what the "values" of this multivalent feature should be. In a full treatment, the reasonable thing to do would be to split the feature into subdivisions. However, given the particular problem under investigation, this seems overmeticulous. The point is to set off the CHS pattern against other types, which leads to a lumping solution. The CHS pattern is for the article to occur once only, on the Dept;¹⁷ if the CHS pattern is

¹⁷ It would seem that the pattern as found in Celtic and Hamito-Semitic in fact specifically demands a medial article: "Head the-Dept". In Aramaic, however, where the article is suffixal, the pattern is "Head Dept-the". The common denominator is that the article, whether prefixal or suffixal, must occur only on the Dependent. One could further insist on the order "Head-Dept", excluding "Dept-Head". But this would effectively debar most OV languages from consideration, something which seems to me undesirable for multiple reasons --- qualitatively (a "Dept-the Head" rule would be highly interesting), quantitatively (the phenomenon is extremely rare on any formulation), and theoretically (it would reduce the independence of the features, making this feature dependent on feature 2).

required (or dominant) in the language, score +1 (or +1/2, respectively). If it is normal for both Head and Dependent simultaneously to take an article, score -1; likewise if only the Head normally takes the article, something which appears to be quite rare (Fijian, as an option). As intended, this has the effect of lumping two quite distinct patterns.

d) If only one article appears possible in the genitive construction, but without restriction as to Head or Dept, score 0 (true zero). This may happen especially if the article is constrained to fall at the edge of the N-Gen syntagm (Wolof; cf. the //nāā demonstrative in Hottentot).

e) Where definiteness is indicated purely as a suprasegmental on the noun itself, as in various African languages (tone change), any questions of relative positioning of article and noun are moot; the "article", if it can be said to exist at all, is coterminous with the noun. But the central issue of localization of definiteness marking --- Head, Dept, or both --- still makes perfectly good sense. I will count such languages as having articles, and score them the same as any other language with articles.

f) Several languages require a possessive pronominal copy on the head noun of N-Gen embeddings, yielding constructions like

(the) man (the) his-house.

Crosslinguistically, there are languages that exclude the copresence of an article and a possessive pronoun with the same noun (Abkhaz, English); but more commonly the two can freely cooccur. The presence of a possessive pronoun in the genitive construction, therefore, should not be taken as fundamentally altering the potential possibilities for article placement: in principle the article might occur on Head or Dept or

both, just as with languages that do not require the possessive.

Accordingly, no special scoring provisos will apply to this subtype.

g) "Hedged articles": Many languages lacking clearly profiled articles do have a "previous-mention" demonstrative element which the grammar presents as verging on articlehood. Typically such demonstratives are common but not obligatory in definite contexts. Should these be reanalyzed and taken as articles? The compromise position I adopt here is to reanalyze some such cases as "hedged articles", yielding scores of $\pm 1/2$, never ± 1 , for feature [8]. In deciding when to reanalyze, I took a case-by-case approach, with attention to various factors:

(i) Frequency of occurrence, either in texts (if any were given) or in examples in the grammar book itself: an article should surely be quite a bit more frequent than a demonstrative. Frequent occurrence with the HeadN of a relative clause was a point in favor of reanalysis (Burushaski, Slave).

(ii) Formal status: reanalysis as an Art was more likely if the element was characterized as basically adnominal, less likely if basically pronominal. An element presented as a noun affix stood an especially good chance of reanalysis (Maricopa). A demonstrative which showed unique patterning, different from other demonstratives, was likewise a good candidate (Lahu).

(iii) Obligatoriness or optionality in given semantic context (previous mention): obligatory occurrence strongly favored reanalysis (Babungo).

Assessments of (i), (ii), (iii) were inevitably impressionistic; most of the time the question could not have a "right" answer. I usually chose to err on the side of caution, not counting the element as an article

unless there were strong grounds for doing so (see individual language write-ups in Appendix 2). These are the languages I did reanalyze as having a "hedged article": Babungo, Burushaski, Cree, Lahu, Maricopa, Slave. --- In practical terms, fortunately, the question was less troublesome than might at first have appeared. Its answer affects only feature [8]; and it happened that for many of the relevant languages, the grammar book presented no genitival examples that included the "article". Such languages would thus end up with a zero score (of two different kinds) on either approach: "I" (Irrelevant) if the particle is not taken to be an article, "N" (No data) if it is.

h) If the definiteness marker does not occur on the noun itself but only on modifiers (e.g. Adjectives, as in Nkore-Kiga), again treat as a "hedged article" with the same scoring restriction.

i) The article in Persian can occur only on the Obj; score -1/2 (like a "hedged article").

j) Bibliographical notes: Chafe 1976:38ff. discusses the concept of definiteness per se; Krámský 1972 presents theoretical discussion and a typological survey; Moravcsik 1969 (esp. 85ff.) discusses articles crosslinguistically within the broader context of determination; Greenberg 1978 discusses the evolution of definite articles to gender markers. None of these treats the positioning of articles in genitive embeddings.

9) Nonconcord with explicit (noun) subject

a) For +1 score, the verb must assume a maximally unmarked form when an explicit full-NP subject is present (perhaps with further

restrictions). Specifically, either (i) or (ii) must hold as a dominant pattern:

(i) Verb assumes 3-sg form with any explicit 3-pers Subj (or perhaps even with non-3-pers independent pronoun Subj), possibly with word order constraints. This is chiefly a matter of number nonconcord.

(ii) Verb would normally take a marker for pers/number of Subj, but the marker disappears if an explicit full-NP Subj is present, subject to word order constraints. Number is irrelevant here.

b) Regarding word order constraints: the CHS pattern is to have nonconcord when a full-NP subject follows the verb. However, in Yagua (VSO) nonconcord occurs precisely when a full-NP Subj precedes the verb; in a few other languages nonconcord involves no stated word order restriction. Overall, the phenomenon was rare enough so that it seemed reasonable to treat all these as variants on a theme, and score all +1.

c) If the verb never shows any Subj marking, score zero ("I").

d) As regards number nonconcord, I concentrated in most languages on ordinary plural nouns displaying overt morphological coding for plural. In some languages having the category number, however, number coding on nouns is nonexistent (Fijian) or rare (e.g. Amele, Koasati, Maricopa, Tauya). Here notional plurality of Subj (or Obj) NPs is typically indicated just on the verb, as the grammar book often stated explicitly. To compensate for the morphological paucity of plural nouns in such languages, I gave special attention to concord patterns involving conjoined NPs and counted NPs (N+Numeral), since here the notion "plural Subj NP" can be overtly indicated through the syntax. Nonetheless, given the overwhelmingly verb-centered expression of plurality in such languages, the IE notion of "concord" is relatively unimportant. I

assigned a score of -1 or -1/2, depending on whether the language showed any evidence at all of CHS-style nonconcord.

e) Some languages which do show full number concord with ordinary plural NPs may allow nonconcord with conjoined NPs or counted NPs (Hungarian, Nubian). If either of these holds as a dominant pattern, score +1/2; if merely as an option, -1/2.

f) Some languages are very flexible regarding number concord between Subj and verb: thus, in contexts where Subj is conceptually plural, plurality might be optionally marked on N or V or both (or neither). Score such languages +1/2 (Chumash, Gilyak, Slave, Tagalog).

(g) Many languages show nonconcord controlled by animacy (inanimates, irrational beings, neuter gender: thus in several IE languages, Georgian, Mangarayi, Turkish); but this has little to do with the phenomenon seen in the CHS languages. Hence if animacy is the only (or major) relevant factor, with number nonconcord holding only (or chiefly) with inanimate nouns, score -1/2.

h) In some languages, a concord prefix on the verb sometimes disappears if preceded by a full NP (Abkhaz, Hixkaryana); but as the phenomenon only occurs under particular phonological circumstances, the process arguably has more to do with low-level morphophonological adjustment and less to do with nonconcord. Score -1/2.

i) I ignore all manifestations of nonconcord that do not directly involve the verb, e.g., the nonconcord between plural HeadN and obligatorily singular resumptive pronoun seen in Subject RCIs in Yoruba (an isolating language). The plural HeadN implies a semantically plural verb in the RC1 (e.g. "the men who know..."), so that the appearance of a singular Subj pronoun in the RC1 might indeed seem an instance of

Subj-Verb nonconcord. It will not, however, be so counted.

10) Verbal abstract: verbal noun (VN) or infinitive (Inf)?

a) This feature imposes a metarequirement: verbs must have a nonfinite nominal verbal abstract (indicating the doing of the verbal action) as a more or less standard part of the verbal paradigm, distinct from a nominalized finite clause and also distinct from a concrete noun --- that is, "slapping" or "to slap", but neither "that he slapped" nor "a slap". Not all languages meet this requirement. The diagnostic that is used to distinguish verbal abstracts from nominalized finite clauses (or indeed from finite verbs) will be tailored to the type of language:¹⁸

--- In a language where the finite verb morphologically codes Subj, the verbal abstract should normally not do so, or only optionally (though it may code Obj). It is useful to allow (not require) this criterion to have a special ergative version. A reasonable criterion for finiteness in some ergative languages is that the verb should code S or O;¹⁹ if so, a nonfinite form should fail to do so. This is the ergative analogue to insisting on Subj coding in a nominative/accusative language. (More discussion under feature [15] below, re Eskimo, Abkhaz.)

¹⁸ Recall the discussion of finiteness in sec. 4.3.

¹⁹ I use the now-standard terminology of Dixon (1979:61):

A = Transitive subject
S = Intransitive subject
O = Transitive object

Ergative languages treat S and O similarly, with A the "odd man out"; in "accusative" languages, it is S and A that are treated the same.

--- In a language where the finite verb itself codes no arguments, the verbal abstract should lack some salient marker normally present on finite verbs (e.g. tense or mood). If this last criterion yields a "verbal abstract" which is almost but not quite identical to the "finite" verb (Dyirbal), treat the metarequirement as weakly satisfied: this will be expressed by assigning scores of $\pm 1/2$ (rather than ± 1).

--- If even this criterion fails, score zero ("I"): the language does not have a category "verbal abstract" distinct from finite verbs. Note that this zero score can occur either because the language is isolating, or because it always uses fully finite (i.e. subject-coded) verbs even in subordinating contexts.

b) In each language an attempt was made to verify that a given "infinitive" or "verbal noun" indeed functioned in prototypical infinitival/gerundial contexts: as complement to verbs (especially EQUI verbs), in purpose clauses, as complement to AUX, as embedded clausal subject, etc. But this was not done rigorously, for both practical and conceptual reasons. On the one hand, grammar books were not always clear on the range of functions a given nonfinite form might assume. On the other, it was not clear to me exactly which usages properly ought and ought not to count as "prototypical infinitival/gerundial" behavior. I did try to ensure (i) that the candidate form be truly nonfinite, (ii) that it not be primarily participial (adnominal, adverbial) in function, and (iii) that it not be a special form used only for clause chaining or serial verbs. Within the context of feature [10], I took note of principles and clause-chaining or serial-verb forms only if they were the

only nonfinite forms in the language.

c) Given the existence of a verbal abstract, the point at issue is to distinguish its two subtypes (VN, Inf). This distinction is meaningful only for transitive verbs. For a VN, the notional object must appear as a genitive (however that is marked). For an Infinitive, the notional object must appear in whatever "case" it would assume for the corresponding finite verb --- prototypically the accusative, which will be used hereafter as a kind of shorthand for "finite verb rection". But if the finite verb lexically governs some other case, the Inf will follow suit. And for ergative languages the "object" of an Inf, like that of a finite verb, will standardly occur in the Absolutive (= Nominative). --- Note that nothing is said here about the case of the Subj of the verbal abstract (if indeed the Subj appears at all).

d) Usually the language will provide some clearcut structural difference between genitive and accusative rection, and thus between VN (+1) and Inf (-1). If the language provides no formal criteria for distinguishing accusative and genitive rection, the feature value is indeterminate: score 0 (a true "0" score), as in Gilyak.

e) When the language has one clear, dominant pattern for deriving verbal abstracts productively from verbs, I will focus on that pattern and ignore minor patterns. Note that languages with productive Infinitives may have lexically governed Verbal Nouns as a minor pattern; here I will focus on the Inf as the dominant pattern. But if VN and Inf can both be formed productively, score 0 (true "0"), as in Hungarian or Maasai.

f) Some languages that lack obvious verbal abstracts (Fijian, Squamish, probably Maricopa) have a finite clause nominalization strategy with a twist: the nominalized verb preserves its finite arguments and

essentially all of its finite categories, but recasts the Subj (not Obj) as a genitive. Should this count in any sense as a "verbal abstract"? The genitive case of the Subj might be taken as adding a degree of noun-iness to the verb; everything else speaks for finiteness. I decided in general to treat such forms not as verbal abstracts but as nominalized finite clauses, rationalizing the genitive as reflecting a voice change of some kind; hence score zero ("I").

g) Hixkaryana has a clause nominalization strategy with person/number and tense still coded, but with all categories (including tense) marked just as they would be in normal nominal possession (quite different from finite verb marking). This will count as a verbal abstract (in fact as VN).

11) Predicative particle

a) The ideal is that a special-purpose predicative particle should exist for use in copular or nominal sentences involving Predicate Nouns or Adjs. Score +1 if the particle is identical (or nearly so) to an Adposition, +1/2 otherwise. The grammar may speak of a predicative suffix or clitic rather than a predicative particle; take all these on a par.

b) It sometimes happens that there is a predicative or declarative particle which can function not only in nominal (copular) sentences but also in true verbal sentences (Chumash, Gilyak, Hua, Lahu, Tauya); score 0 (true 0). And there are languages where a predicative particle has additional particulate (i.e. non-Adpositional) uses that have nothing to do with predication (Afar, Chrau, Hawaiian, Quechua); score -1/2. The

rationale is to set up a semantic gradient involving various degrees of "hedged" predicative particles: having an extra function which is specifically predicative (Chumash etc.) seems closer to the ideal than having various random extra functions (Afar etc.), and the scoring should reflect this.

c) The considerations in (a), (b) show that the "predicative particle" feature, despite initial appearances, is not naturally bivalent. The above intermediate scores represent an attempt to reduce a multivalent feature to a bivalent one by linearization.

d) There may sometimes be confusion in categorizing a predicative element X as particle or as copular verb. Most of the time the choice is straightforward. X is clearly a copular verb if it inflects like a verb. It is normally a particle if, e.g.,

- (i) it cooccurs with a copular verb
- (ii) it occurs instead of a verb in equational sentences,
but does cooccur with verbs in other sentence types
- (iii) it is noninflecting (in a language where verbs inflect)
- (iv) it is affixal rather than a separate word (a weak criterion)
- (v) it is identical to a preposition.

The difficulty arises in cases where there is no straightforward copular verb, and where X (occurring in lieu of copula) does inflect, but not like a verb. This is the situation in several Afroasiatic languages (Semitic, Egyptian, Hausa), where the predicative element X resembles or is identical to an article, demonstrative, or 3rd-person pronoun, coding gender and number. (In Hausa, further, the element X does not occupy the same slot as a normal verb: the language is SVO, yet the predicative

syntagm is Subj PredN X.) To count an inflected element as a "particle" would seem to strain the normal sense of the term; hence in all these cases, the element X will be counted as a special kind of copular quasi-verb, and not as a predicative particle. Score -1.

e) If there are two (or more) distinct predicative constructions, one with particle and one without, score +1 if the particle construction is of major importance, +1/2 if secondary.

12) Adpositional periphrastic

a) The intended construction (score +1) is

Copula + Prep + some nonfinite form ["BE at VERBing"],

or its Postpositional equivalent; we will expect adpositions with meanings like "on, in, at". The construction should represent a regular "periphrastic inflection" of the verb. For this feature, the notion "nonfinite form" is construed more broadly than in [10]. In addition to well-profiled verbal abstracts it will include other nonfinite forms, and will admit the all-purpose uninflecting form typically found in an isolating language like Chinese, a form which is intrinsically neither finite nor nonfinite (since no meaningful opposition exists). An isolating language will thus not have an automatic "I" score. However, if a language makes exclusive use of explicitly finite verb forms, score zero ("I"). --- If the language has only "hedged adpositions", scores may range only between +1/2 and -1/2. If the language lacks the category Adposition, score zero ("I"). NB: These last two provisos are conceptually irrelevant if one of the nonadpositional variants mentioned under (d) below is involved, and will not apply in such cases.

b) The construction should normally have either neutral or (roughly) progressive/durative semantics; if more specialized semantics, score +1/2 (Hixkaryana). If the Prep is "without", score -1/2 (Albanian); the implied privative meaning is very far from the prototype. NB: Specifically excluded here is the Prep "to" with purposive or future or obligational semantics (as in English "is to VERB"); this will score -1.

c) Some languages (Celtic, Egyptian) allow several different adpositions in this construction, each conveying different tense/aspect information. This multiple-Adp type will be scored the same as the commoner single-Adp type, with priority given to the present or progressive variety (if any).

d) The construction has several minor variants, all scored +1/2:

--- Instead of an Adp, a non-Adpositional particle with clear Adp-like meaning occurs.

--- No Adp occurs, but the nonfinite form is in a "local" case (e.g. Locative).

--- Neither Adp nor local case occurs, but the Copula is specifically a locative copula ("to-be-at") (Lahu). Somewhat similar is Abkhaz, with the PostpPhr "in it" recast as a verb (to "in-it").

e) Note that these +1/2 scores (in [d]) represent a way of reducing a multivalent feature to bivalency by linearization.

f) Tzutujil has several constructions with progressive semantics, but involving Aux verbs "DO" and "be-in-the-act-of" rather than "BE". These constructions seem conceptually to straddle [12] and [13], perhaps somewhat closer to [13]. Compromise by scoring [12] as zero and [13] as +1/2. (See also [13b] below.)

g) For an examination of progressive constructions

crosslinguistically, see Blansitt 1975, esp. pp. 14ff.

13) "DO" Periphrastic

a) The intended construction (score +1) is

DO + some nonfinite form ["DO VERBing"]

or its OV equivalent, where the nonfinite form can be derived productively from all verbs or a large class subclass of verbs. The construction will thus represent a regular "periphrastic inflection" of the verb. As with feature [12], the notion "nonfinite form" will be broadly construed to admit not just verbal abstracts but nonfinite forms of any sort, as well as the uninflecting form of isolating languages. Isolating languages will thus not score "I". Languages having only explicitly finite verb forms will score zero ("I") unless the syntagm involves the incorporation of a verb root (Kiowa) or the nonverbal construction in (c) below.

b) Semantically, the construction should either have emphatic or neutral semantics; if more specialized semantics (Hua, Lahu), including the case of progressive semantics (Kiowa, Tzutujil), score +1/2. Specifically excluded is the very common case where DO + VN expresses the Causative; score -1 (unless some other use exists too).

c) It is very common, when a language borrows foreign verbs, for it to recast them as nominal Stems; the new verb then exists in the recipient language only in periphrastic form (DO Stem), with only the Aux verb (DO) inflecting. This defines a subtype distinct from the prototypical construction (above), for Stem has no lexical link to a verb in the recipient language. "Compound" locutions of this type can abound in a

language, often representing a productive strategy for borrowing. And the construction can exist with native lexemes too, with the nonfinite Stem either drawn from nonverbal lexemes (N, Adj, Adv, etc.) or lexically "isolated" and found only in this syntagm. (An unusual type is "DO Ideophone", very common in Hixkaryana.) In isolated cases, but only in isolated cases, Stem may even be explicitly deverbal. But in this type Stems should not show any productive link to verbs. --- The essential feature of this subtype is that it is lexical, not grammatical. It matches our intended construction syntagmatically, but not paradigmatically or functionally; in contrast to English "DO Verb" ("he does sing"), for example, or Welsh "DO Verb-ing", it does not coexist paradigmatically alongside the simplex form ("he sings") as a "periphrastic inflection" of the verb. Rather, it is a complex word-formation device. Such cases will be scored zero (true 0).

d) Some languages have a construction which is exactly right except that the Aux is not DO but some other verb ("say", "put", "hit", etc.). Score such cases +1/2 (Afar, Amele).

e) If the language has serial verb constructions, such that verbs regularly form complex (serial) nuclei with DO, score +1/2 (Yimas).

14) Adverbial clause = "and" + finite clause

a) Conceptually, this was probably the most ill-defined feature in the feature set, for there were no straightforward criteria for saying when the syntagm "and + Finite Clause" (hereafter A+FnCl) was really being used as a notional adverbial clause --- nor what a "notional adverbial clause" actually was. It is, after all, routine for "and" conjunction

in many languages (e.g. English) to cover a wide range of possible implicit temporal nuances, yet without qualifying as an "adverbial construction". Not infrequently, grammar books would indeed suggest translations for A+FnCl in adverbial terms ("S1 and S2" might be translated "S1 while S2"), or would explicitly state that A+FnCl had particular adverbial senses. But rarely was it clear whether these translations and explanations were meant to reflect an actual structural reality within the language, or were merely expository heuristics.²⁰ In spite of this, I had little choice but to rely heavily on such statements; when they appeared in the grammar book, I took them as reflecting something real about the language.

b) If the grammar book explicitly presents clear adverbial uses for the construction, or renders it with adverbial translation equivalents in English, score +1 or +1/2 depending on the semantics (see below); if no such use is mentioned, score -1. Verbless "Absolute" constructions with "and" ("and me unwilling to wait") will score +1 if the elements accompanying "and" can stand alone as an independent sentence in the language, -1 otherwise.²¹ If the adverbial nature of the clause seems

²⁰ Within the grammatical traditions of Celtic and Semitic, the notion of A+FnCl as constituting a real adverbial clause construction has always been accepted as natural and intuitive, and my own experience with the languages confirms this strongly. Significantly, there is usually in these languages a secondary formal mark on A+FnCl: either the clause is nominal (verbless), or its verb assumes an imperfective form, or the normal VS order reverses to SV, all devices which function in the languages to signal backgrounded (adverbial) information. One might perhaps examine the languages in the sample for such secondary formal marks; but relevant information was seldom forthcoming from the grammars.

²¹ Note the differences between this and the participial "Absolute" construction of Indo-European ("John having arrived, I left"; see Chapter 3, discussion of feature [15]). That construction involves an explicitly nonfinite verb, assigns no crucial role to the presence of "and", and cannot stand alone as an independent sentence.

more than usually in doubt, score +1/2. Considerable subjectivity is unavoidable here. If the language has no sentence-level "and" conjunction, score zero ("I").

c) The prototypical phenomenon as found in the CHS languages involves a range of adverbial nuances centering on "concomitant circumstance", and it seemed reasonable for the scoring to be responsive to this. Accordingly, a language will be scored +1 only if the adverbial nuances described for the "and"-construction include the semantic range of "when, while, just as, etc.", or are presented in very broad and inclusive semantic terms. If, on the other hand, the semantics involve only such nuances as cause, purpose, or conditionality, score +1/2.

d) The term "finite" will be construed broadly, so as to include the uninflecting form of isolating languages. (Compare [12] and [13], where nonfinite is broadly construed, in identical fashion.)

e) In some languages (Lahu), the grammar mentions an adverbial use for the construction "Clause1 and Clause2", but with polarities reversed --- that is, the notional adverbial is "Clause1 and", while Clause2 is the notional main clause. This seems a natural OV analogue to the construction as presented above, and to exclude it purely on the basis of word order would seem arbitrary (recall the discussion under feature [8] above). To be sure, this construction is not always easy to distinguish from clause chaining (feature [15]); for our purposes, however, the requirement that the clause be finite and that an explicit "and" conjunction be present will keep the two apart.

f) Bibliographical note: Talmy 1978 investigates the use of "and"-coordination for notional subordination.

15) VN/Inf instead of finite main-clause form

a) The ideal here (score +1) is a construction where a VN/Infinitive occurs (instead of a finite form) for the equivalent of main-clause predication. The phenomenon occurs par excellence in clause chaining or the equivalent (recall the discussion in sec. 4.3), and in most grammar books was presented under this heading.²² --- Note, however, that in some languages there is a special nonfinite form used specifically for clause chaining, one which has no uses plausibly falling under the header of "VN/Inf" (feature [10]). All other things being equal, such cases will be scored +1/2, as compared to languages where the clause-chaining form is itself a VN/Inf (score +1). The scoring decision here depends on whether the special clause-chaining form can in some sense "count as a VN/Inf", that is, can fulfill prototypical VN/Inf functions (e.g. "purpose clauses") --- a judgment which, unfortunately, is prone to both subjectivity and error.

b) Here (unlike features [12], [13]) I will insist on true nonfiniteness, requiring that the language have a clear opposition between finite and nonfinite forms; if not, score zero ("I"). Unlike [12] and [13], the whole point of this feature is specifically to distinguish finite and nonfinite, so there is no point in counting languages which lack the distinction.

c) Some ergative languages have clause-chaining forms that appear to

²² Given the constrained notion of nonfiniteness used in this study, I will ignore the finite (i.e., subject-coded) clause-chaining forms which are found in many languages (e.g. Amharic, many New Guinea languages). This includes such forms as the subject-coded Consecutive tenses found in languages like Swahili (the -ka- form) and Coptic (Conjunctive).

be finite except for a single peculiarity: they omit Subj coding on transitive verbs (Eskimo, Abkhaz). Failure to code Subj is normally our primary criterion for nonfiniteness. Given that the language is ergative, however, such forms will be taken to be fully finite, inasmuch as they do code S and O, a natural ergative constellation analogous to "Subject".²³ Being finite, these forms will be ignored here. (Contrast the "real" verbal abstracts in these languages, which are morphologically much further away from finite verbs.)

d) If a real or apparent "nonfinite" form functions only within serial verb constructions, the phenomenon should not count as clause chaining: not clauses but verb stems are being joined, and the form does not truly stand in for a main-clause predication. Score -1.

e) If the only main-clause use of nonfinite forms is in wishes, commands, etc., score -1/2 (Greek, Amele).

f) Bibliographical note: Longacre 1990 presents a series of case studies of clause chaining as used in storytelling in numerous African languages; in many of the VO languages he discusses, clause-chaining phenomena appear where the verb that takes full finite marking is the first verb in the chained sequence. This is similar to the CHS situation, and differs from prototypical clause chaining (as seen, for example, in New Guinea), where the verb that gets full finite marking is the final one in the sequence.

²³ For the symbols S, O, A, see note to feature [10] above.

16) Word-initial change

a) If the language has syntactically controlled word-initial change as a clear and dominant part of its grammar, score +1. Since the phenomenon is quite rare, I lumped together (i) the type where the change codes a single grammatical category and (ii) the multicategorical type.

b) If the phenomenon is restricted to a relatively small corner of the grammar, or has strong lexical constraints, score +1/2. If it represents only lexical traces of a once-productive phenomenon, score -1/2 (Lahu). If a change shows up initially in some words but noninitially in others (Basque), score -1/2.

c) In several languages (Bantu, Ingush), person/number or noun-class affiliation is coded at the beginning of the word. Because these markers express agreement, this is technically an instance of syntactically controlled word-initial change; but the phenomenon is as much lexical as syntactic, and clearly differs profoundly from what happens in Berber or Celtic. Score -1.

d) If the grammar omits all mention of the topic, assume nonoccurrence (score -1).

e) For crosslinguistic discussion of initial change see Ultan 1970, Lieber 1987 [Chapter 2: Mutation]; for West Africa, see Sapir 1971.

17) Extended use of kin terms

a) The construction in question will ideally be a clearcut genitive "KIN of Noun", in a nonliteral usage paraphrasable as "s/he of Noun, it

of Noun"; the meaning should be "someone/something which is characterized by some kind of essential connection to Noun", e.g. "son of sweat" = "laborer" (Tagalog). The semantics are important. It is not uncommon, for example, to find languages where kin terms function as augmentatives or diminutives (Matisoff 1992), but these will not count as examples of our construction; nor will the numerous unanalyzed compounds involving -zi (originally "child", now semantically empty) in Chinese.

b) If no lexical data is available, score zero (code "N"). For a language whose source is a Croom-Helm (or Lingua Descriptive) grammar, the mini-lexicon at the back of the grammar will be deemed a suitable source of lexical data, in the absence of any more comprehensive dictionary; such languages will be given nonzero scores.

c) Standard procedure is to check the lexicon under words for "father", "mother", "son", "daughter", and "child" (but not others). If numerous good examples can be found, score +1. If numerous good examples exist, but the construction involves not a genitival combination but a compound "KIN-Noun", score +1/2. If the phenomenon is not frequent but the available lexical data include at least 4 good examples, score +1/2. If the data include 1-3 good examples, consider the phenomenon to exist only in traces and score -1/2. If the data give no indication that the phenomenon exists at all, score -1.

d) This feature will show a systematic scoring bias arising from the crudeness of the lexical materials consulted for most of the languages. Checking in better dictionaries would undoubtedly raise the score in some cases --- as actually happened with Hungarian, where only information gleaned from a large etymological dictionary made it clear that the phenomenon really does exist in the language. Direct work with

informants or experts in the language could also be expected to raise the score (as happened with Wolof).

e) For crosslinguistic discussion of the nonliteral uses of "mother" and other kin terms, especially in Southeast Asia, see Matisoff 1992. The focus in that article, however, is on the use of the "mothermorphs" in augmentatives and diminutives, not on the kind of metaphorical usage which is at issue here.

Chapter 6: Results and analysis

6.1 The raw data

The numerical results of the survey are presented in full in Table 1a. The numbers are compressed slightly in Table 1b, with the relative clause feature-pair 4aA (similarly 4bB) combined and averaged as described in Chapter 5 so as to yield three (not five) RC1 features, viz. 4a, 4b, 4c. This will serve as the master data table for the remainder of this study.

Key to Table 1b (Features)

1	Conjugated Adpos	8	Def article with genitives
2a	Word Order: clause	9	V nonconcord with NP Subj
2b	Word Order: NP	10	Verbal noun or infinitive
3	RC1: linker	11	Predicative particle
4a	RC1, Genitive: Gap/Copy	12	Adpositional periphrastic
4b	RC1, Adpos: Gap/Copy	13	"DO" periphrastic
4c	RC1, Adpos: Move Adpos	14	Adv clause = AND + finite cl
5	RC1: special verb form	15	VN/Inf instead of finite cl
6	Polypersonal verb	16	Word-initial change
7	Infix/suffix alternation	17	Extended use of kin terms

	1	2a	2b	3	4a	4A	4b	4B	4c	5	6	7	8	9	10	11	12	13	14	15	16	17
NE Abkhaz	+1	-1	-1/2	+1	-1	+1	-1	-1	+1	+1	+1	-1	+1	-1/2	+1	+1/2	+1/2	-1	-1/2	-1	-1	-1
AF Afar	-1/2	-1	-1/2	+1			-1	-1	+1/2	+1/2	-1			+1	-1	-1/2	+1	+1/2	-1	-1	-1	N
NE Akkadian	+1	-1	+1	+1	+1	+1	+1	+1	-1	0	+1	-1		-1	+1/2	-1	-1	-1	+1	-1/2	-1	+1/2
EU Albanian	-1	+1/2	+1	0	-1	-1	-1	-1	-1	-1	+1	+1	-1	-1	-1	-1	-1/2	-1	+1/2	-1	-1	-1
NG Amele	-1	-1	0	+1	+1/2	+1/2	-1/2	-1/2	+1/2	-1	+1	-1		-1	-1	-1	-1	+1/2		-1/2	-1	-1
AF Amharic	+1	-1	-1	+1	+1	+1	+1/2	+1/2	+1/2	0	+1	-1/2	+1	-1	-1	-1	+1	+1/2	+1/2	-1	-1	+1/2
NE Arabic (Cl)	+1	+1	+1	+1	+1	+1	+1	+1	-1	-1	+1	-1	+1	+1	+1	+1/2	-1	-1	+1	-1	-1	+1
NE Arabic (Mod)	+1	+1/2	+1	+1	+1	+1	+1	+1	-1	-1	+1	-1	+1	-1	+1/2	-1	+1	-1	-1	+1	-1	+1
AF Babungo	-1	+1/2	+1	+1	+1	+1	+1	+1	-1	-1	-1	-1		N	-1	-1	-1	-1		-1	-1	+1/2
EU Basque	-1/2	-1/2	-1/2	+1			-1/2	-1/2	-1/2	0	+1	-1/2	-1	-1	-1	-1	+1/2	+1	+1/2	+1/2	-1/2	-1
AF Berber	+1	+1	+1	+1			-1	-1	+1	+1/2	+1	+1		-1		+1					+1	+1/2
EA Burushaski	+1	-1	-1	+1	+1	+1				-1	+1	0	-1/2	-1	-1	-1	-1	0	+1/2	+1/2	-1	-1
EA Chinese	-1	+1/2	-1/2	+1	+1/2	+1/2	-1/2	-1/2	-1	-1	-1					-1	+1	-1		-1	+1/2	
EA Chrau	-1	+1/2	+1	+1						-1	-1					-1/2	-1	-1			-1	N
NA Chumash		+1	+1/2	+1	+1	+1				+1/2	+1	-1	-1	+1/2		0			+1	-1	-1	-1
AF Coptic	+1	+1/2	+1	+1	+1	+1	+1	+1	-1	-1	+1	-1	-1	+1/2	-1	+1	-1	0	-1	-1	-1	-1
NA Cree		+1	0							0	+1	+1/2	-1/2	-1		-1		-1		-1/2	-1	
NG Dyirbal				+1						+1	-1		-1	-1/2	-1		-1		+1/2	-1	N	
AF Egyptian	+1	+1	+1	+1	+1	+1	+1	+1	-1	+1/2	-1/2	-1		+1	+1	+1	+1	+1		+1	-1	-1
EU English	-1	+1/2	+1/2	0	-1	-1	-1	-1	-1/2	-1	-1		+1/2	-1	-1/2	-1	-1	+1	-1	-1	-1	-1
NA Eskimo	+1/2	-1	-1/2	+1	-1	+1				0	+1	-1/2		-1	+1/2	+1/2	-1/2	-1	+1	-1	-1	-1
OC Fijian	-1/2	+1	+1	+1	N	N	-1	-1	+1/2	-1	-1		-1	-1/2		-1				-1	-1/2	
EU French	-1	+1	+1	-1	-1	-1	-1	-1	-1/2	-1	+1	+1	-1	-1	-1	-1	+1/2	-1	-1	-1	-1	-1
AF Gbeya	0	+1/2	+1/2	+1	+1/2	+1/2	+1/2	+1/2	+1/2	-1	-1	-1	0			+1	-1	-1	+1/2	-1	-1	+1/2
AF Geez	+1	+1	+1/2	+1	+1	+1	+1	+1	+1/2	-1	+1	-1		-1/2	+1/2	-1	-1	-1	+1/2	-1	-1	+1/2
NE Georgian	-1	-1/2	-1/2	-1	-1	-1	-1	-1	-1	-1	+1	-1		-1/2	+1	-1	-1	-1	-1	-1	-1	-1/2
EA Gilyak	+1/2	-1	-1	+1						+1	-1/2	-1		+1/2	0	0	-1	-1		+1/2	+1	-1
EU Greek (Cl)	-1	-1/2	+1/2	-1	-1	-1	-1	-1	-1/2	-1	-1		-1	-1/2	-1	-1	-1	-1	+1	-1/2	-1	-1
AF Hausa	+1/2	+1/2	+1/2	+1	+1	+1	0	0	-1	+1/2	+1	-1	+1	-1	+1/2	-1	+1/2	-1		-1	-1	+1
OC Hawaiian	-1	+1	+1/2	+1			-1	-1	+1/2	+1/2	-1		-1	-1	-1/2	-1	-1	-1	+1	-1	-1	-1
NE Hebrew	+1	+1	+1	+1	+1	+1	+1	+1	-1	-1	-1	-1	+1	+1	+1/2	-1	-1	-1	-1	+1	+1	+1
NE Hittite	+1	-1	-1		+1	-1	+1	-1	-1	-1	-1			-1/2	-1/2	-1	-1	-1	-1	-1	-1	-1
SA Hixkaryana	+1	-1/2	-1							+1	-1/2		-1/2	+1	+1	+1/2	0		-1	-1/2	-1	
AF Hottentot	-1	-1/2	-1	+1	+1/2	+1/2	+1/2	+1/2	-1	0	+1	-1		-1	-1/2	-1	-1	-1/2	+1/2	-1	-1	-1
NG Hua		-1	-1	+1	-1	+1				0	+1	0		-1	-1	0		+1/2	+1/2	-1	-1	+1/2
EU Hungarian	+1/2		-1/2	-1	-1	+1	-1	-1	-1/2	-1	0	-1/2	0	-1/2	0	-1	-1/2	-1	-1	-1	-1	+1/2
NE Ingush	-1	-1/2	-1	+1	-1	-1	-1/2	-1/2	-1	+1/2	-1/2	-1		-1	-1		0			-1	-1	-1
EU Irish (Old)	+1	+1	+1	+1/2	+1	+1	-1/2	-1/2	+1	+1/2	+1	+1	+1	-1	+1	-1	+1	-1	+1	-1/2	+1	+1
EU Irish (Mod)	+1	+1	+1	+1/2	+1	+1	+1	+1	-1	+1	-1		+1	+1	+1/2	+1/2	+1	-1	+1	-1	+1	+1
EA Japanese	-1	-1	-1	+1	+1/2	+1/2	-1/2	-1/2	-1	-1/2	-1			-1	-1	+1/2	0	-1/2	+1	-1	-1	-1
NA Kiowa	-1/2	-1/2		+1/2	-1/2					-1	+1	-1/2		-1		-1		+1/2	+1/2		-1	N
NA Kotsati	+1/2	-1	-1/2	+1						+1	+1	-1	0	-1	-1	-1	-1/2	-1		+1/2	-1	N
NG Kobon	-1	-1	-1/2	+1	0	-1	0	-1	-1	-1/2	-1			-1	-1	+1/2	-1	+1	-1	+1/2	-1	-1
EA Lahu	-1	-1	-1/2	+1						-1	-1		-1/2			0	+1/2	+1/2	+1/2		-1/2	+1/2
AF Lango	+1	+1/2	+1	+1	+1	+1	+1	+1	-1	-1/2	+1	-1		+1	-1	-1	-1	-1		-1	-1	+1
AF Maasai	-1	+1	+1	+1	-1	-1				+1	0	-1/2	-1	-1	0	-1	-1	-1	-1	+1	-1	N
AF Mandinka	-1	0	-1/2		+1	-1	+1	-1	-1	-1	-1		-1		-1	+1	+1	+1		+1	-1	-1
NG Mangarayi			+1/2	+1						0	+1	-1		-1/2	+1	-1		0		+1/2	-1	-1

Table 1a: Uncompressed raw scores, full sample (continues next page)

	1	2a	2b	3	4a	4A	4b	4B	4c	5	6	7	8	9	10	11	12	13	14	15	16	17
NA Maricopa		-1	-1/2	+1						+1/2	+1	-1	-1/2	-1/2		-1	0			-1	-1	
NA Miwok (Lake)		-1	-1/2	+1						+1	-1					-1	+1/2	+1		-1	-1	
SA Mixtec	+1/2	+1	+1	+1			-1	-1	-1	-1	-1					-1		+1/2		-1	-1	
AF Nkore-Kiga	+1	+1/2	+1	+1/2	+1	+1	-1	-1	+1/2	+1/2	+1	-1	N	-1	-1	-1	-1	-1	-1	-1	-1	
AF Nubian	+1	-1	0	+1						+1/2	-1			-1/2	0	+1/2	-1	0	-1/2	-1	-1	
NE Persian	+1	-1	+1	+1	+1	+1	+1	+1	-1	-1	+1	-1	-1/2	-1/2	+1	-1	-1	+1	-1/2	+1/2	-1	-1
SA Pipil	+1/2	+1	+1/2	+1						-1	+1	-1	-1	-1		-1	0	-1		-1	-1	
SA Quechua	-1/2	-1	-1	+1			0	-1	-1/2	0	-1/2	-1		-1	-1	-1/2	-1/2	-1	-1	-1	-1	
NA Shoshone	-1	-1	-1/2	+1/2			+1/2	+1/2	-1/2	0	-1/2	-1/2		-1	-1	-1	-1	-1/2	-1	-1	-1	
NA Slave	+1	-1	-1/2	+1	+1	0	+1	0	-1	-1	+1	-1	+1/2	+1/2		-1		-1		-1	-1	
NA Squamish		+1	+1/2	+1/2	-1/2	+1/2				+1/2	+1	-1/2	-1	+1/2		-1		+1		-1	-1	
NE Sumerian	-1/2	-1	+1/2	+1	+1/2	+1/2	+1/2	+1/2	-1/2	+1/2	+1	+1/2		0	-1	-1	-1/2	+1/2	-1	-1	-1	
OC Tagalog		+1	+1/2	+1						-1	-1				+1/2	-1	-1	-1	-1	-1	+1/2	
EA Tamil	-1	-1	-1	+1	+1/2	-1/2	-1	-1	-1	0	-1			-1	-1	+1	-1/2	0		+1	-1	
NG Tauya		-1	-1/2	+1	-1/2	+1/2				-1	+1	-1		-1	-1	0	-1		+1/2	-1	N	
NE Turkish	+1/2	-1	-1/2	+1	-1	+1	-1	-1	-1	0	-1			-1/2	-1/2	-1	+1/2	0	-1	+1/2	-1	
SA Tzutujil	+1/2	+1	+1/2	+1	-1	+1	-1	0	+1/2	-1	+1	-1	-1/2	-1	0	-1	0	+1/2	-1	-1	-1	
EU Welsh	+1	+1	+1	+1/2	+1	+1	+1	+1	-1	-1/2	+1/2	-1/2	+1	+1	+1	+1	+1	+1	+1	+1	-1	
AF Wolof	-1	+1/2	+1	+1						-1	+1	+1	0	+1	-1	-1	-1	-1	-1	-1	+1	
SA Yagua	+1	+1	0	+1	0	0	0	0	-1	+1/2	-1			+1	-1	-1	-1	-1		-1	N	
NG Yimas	+1		-1/2	+1						+1/2	+1	+1		+1/2	-1	-1	-1	+1/2	-1	+1/2	-1	
AF Yoruba	-1/2	+1/2	+1	+1	+1	+1	-1/2	-1/2	-1/2	-1	-1			-1	-1	-1	+1	-1/2	-1	-1	+1	

1 2a 2b 3 4a 4A 4b 4B 4c 5 6 7 8 9 10 11 12 13 14 15 16 17

Table 1a: Uncompressed raw scores, full sample

	1	2a	2b	3	4a	4b	4c	5	6	7	8	9	10	11	12	13	14	15	16	17
NE Abkhaz	+1	-1	-1/2	+1	0	-1	+1	+1	+1	-1	+1	-1/2	+1	+1/2	+1/2	-1	-1/2	-1	-1	-1
AF Afar	-1/2	-1	-1/2	+1		-1	+1/2	+1/2	-1			+1	-1	-1/2	+1	+1/2	-1	-1	-1	N
NE Akkadian	+1	-1	+1	+1	+1	+1	-1	0	+1	-1		-1	+1/2	-1	-1	-1	+1	-1/2	-1	+1/2
EU Albanian	-1	+1/2	+1	0	-1	-1	-1	-1	+1	+1	-1	-1	-1	-1	-1/2	-1	+1/2	-1	-1	-1
NG Amele	-1	-1	0	+1	+1/2	-1/2	+1/2	-1	+1	-1		-1	-1	-1	-1	+1/2		-1/2	-1	-1
AF Amharic	+1	-1	-1	+1	+1	+1/2	+1/2	0	+1	-1/2	+1	-1	-1	-1	+1	+1/2	+1/2	-1	-1	+1/2
NE Arabic (Cl)	+1	+1	+1	+1	+1	+1	-1	-1	+1	-1	+1	+1	+1	+1/2	-1	-1	+1	-1	-1	+1
NE Arabic (Mod)	+1	+1/2	+1	+1	+1	+1	-1	-1	+1	-1	+1/2	-1	+1	-1	-1	-1	+1	-1	-1	+1
AF Babungo	-1	+1/2	+1	+1	+1	+1	-1	-1	-1		N		-1	-1	-1	-1		-1	-1	+1/2
EU Basque	-1/2	-1/2	-1/2	+1		-1/2	-1/2	0	+1	-1/2	-1	-1	-1	-1	+1/2	+1	+1/2	+1/2	-1/2	-1
AF Berber	+1	+1	+1	+1		-1	+1	+1/2	+1	+1		-1		+1				+1	+1/2	
EA Burushaski	+1	-1	-1	+1	+1			-1	+1	0	-1/2	-1	-1	-1	-1	0	+1/2	+1/2	-1	-1
EA Chinese	-1	+1/2	-1/2	+1	+1/2	-1/2	-1	-1	-1					-1	+1	-1			-1	+1/2
EA Chrau	-1	+1/2	+1	+1				-1	-1					-1/2	-1	-1			-1	N
NA Chumash		+1	+1/2	+1	+1			+1/2	+1	-1	-1	+1/2		0			+1		-1	-1
AF Coptic	+1	+1/2	+1	+1	+1	+1	-1	-1	+1	-1	-1	+1/2	-1	+1	-1	0	-1	-1	-1	-1
NA Cree		+1	0					0	+1	+1/2	-1/2	-1		-1			-1		-1/2	-1
NG Dyirbal				+1				+1	-1		-1		-1/2	-1		-1		+1/2	-1	N
AF Egyptian	+1	+1	+1	+1	+1	+1	-1	+1/2	-1/2	-1		+1	+1	+1	+1	+1		+1	-1	-1
EU English	-1	+1/2	+1/2	0	-1	-1	-1/2	-1	-1		+1/2	-1	-1/2	-1	-1	+1	-1	-1	-1	-1
NA Eskimo	+1/2	-1	-1/2	+1	0			0	+1	-1/2		-1	+1/2	+1/2	-1/2	-1	+1	-1	-1	-1
OC Fijian	-1/2	+1	+1	+1	N	-1	+1/2	-1	-1		-1	-1/2		-1				-1	-1/2	
EU French	-1	+1	+1	-1	-1	-1	-1/2	-1	+1	+1	-1	-1	-1	-1	+1/2	-1	-1	-1	-1	-1
AF Gbeya	0	+1/2	+1/2	+1	+1/2	+1/2	+1/2	-1	-1	-1	0			+1	-1	-1	+1/2		-1	-1/2
AF Geez	+1	+1	+1/2	+1	+1	+1	+1/2	-1	+1	-1		-1/2	+1/2	-1	-1	-1	+1/2	-1	-1	+1/2
NE Georgian	-1	-1/2	-1/2	-1	-1	-1	-1	-1	+1	-1		-1/2	+1	-1	-1	-1	-1	-1	-1	-1/2
EA Gilyak	+1/2	-1	-1	+1				+1	-1/2	-1		+1/2	0	0	-1	-1		+1/2	+1	-1
EU Greek (Cl)	-1	-1/2	+1/2	-1	-1	-1	-1/2	-1	-1		-1	-1/2	-1	-1	-1	-1	+1	-1/2	-1	-1
AF Hausa	+1/2	+1/2	+1/2	+1	+1	0	-1	+1/2	+1	-1	+1	-1	+1/2	-1	+1/2	-1		-1	-1	+1
OC Hawaiian	-1	+1	+1/2	+1		-1	+1/2	+1/2	-1		-1		-1	-1/2	-1	-1	+1	-1	-1	-1
NE Hebrew	+1	+1	+1	+1	+1	+1	-1	-1	+1	-1	+1	+1/2	-1	-1	-1	-1	+1	+1	-1	+1
NE Hittite	+1	-1	-1		0	0	-1	-1	-1			-1/2	-1/2	-1	-1	-1	-1	-1	-1	-1
SA Hixkaryana	+1	-1/2	-1					+1	-1/2			-1/2	+1	+1	+1/2	0		-1	-1/2	-1
AF Hottentot	-1	-1/2	-1	+1	+1/2	+1/2	-1	0	+1	-1		-1	-1/2	-1	-1	-1/2	+1/2	-1	-1	-1
NG Hua		-1	-1	+1	0			0	+1	0		-1	-1	0		+1/2		+1/2	-1	N
EU Hungarian	+1/2		-1/2	-1	0	-1	-1/2	-1	0	-1/2	0	-1/2	0	-1	-1/2	-1	-1	-1	-1	+1/2
NE Ingush	-1	-1/2	-1	+1	-1	-1/2	-1	+1/2	-1/2	-1		-1		-1		0			-1	-1
EU Irish (Old)	+1	+1	+1	+1/2	+1	-1/2	+1	+1/2	+1	+1	+1	-1	+1	-1	+1	-1	+1	-1/2	+1	+1
EU Irish (Mod)	+1	+1	+1	+1/2	+1	+1	-1	+1	-1		+1	+1	+1/2	+1/2	+1	-1	+1	-1	+1	+1
EA Japanese	-1	-1	-1	+1	+1/2	-1/2	-1	-1/2	-1			-1	-1	+1/2	0	-1/2	+1	-1	-1	-1
NA Kiowa		-1/2	-1/2		0			-1	+1	-1/2		-1		-1		+1/2	+1/2		-1	N
NA Koasati	+1/2	-1	-1/2	+1				+1	+1	-1	0	-1	-1	-1	-1/2	-1		+1/2	-1	N
NG Kobon	-1	-1	-1/2	+1	-1/2	-1/2	-1	-1/2	-1			-1	-1	+1/2	-1	+1	-1	+1/2	-1	-1
EA Lahu	-1	-1	-1/2	+1				-1	-1		-1/2			0	+1/2	+1/2	+1/2		-1/2	+1/2
AF Lango	+1	+1/2	+1	+1	+1	+1	-1	-1/2	+1	-1		+1	-1	-1	-1	-1		-1	-1	+1
AF Maasai	-1	+1	+1	+1	-1			+1	0	-1/2	-1	-1	0	-1	-1	-1	-1	+1	-1	N
AF Mandinka	-1	0	-1/2		0	0	-1	-1	-1		-1		-1	+1	+1	+1		+1	-1	-1
NG Mangarayi			+1/2	+1				0	+1	-1		-1/2	+1	-1		0		+1/2	-1	-1

Table 1b: Compressed raw scores, full sample (continues next page)

	1	2a	2b	3	4a	4b	4c	5	6	7	8	9	10	11	12	13	14	15	16	17
NA Maricopa		-1	-½	+1				+½	+1	-1	-½	-½		-1		0			-1	-1
NA Miwok (Lake)		-1	-½	+1				+1	-1					-1		+½	+1		-1	-1
SA Mixtec	+½	+1	+1	+1		-1	-1	-1	-1					-1		+½		-1	-1	
AF Nkore-Kiga	+1	+½	+1	+½	+1	-1	+½	+½	+1	-1	N	-1	-1	-1	-1	-1	-1	-1	-1	-1
AF Nubian	+1	-1	0	+1				+½	-1			-½	0	+½	-1	0	-½	-1	-1	-1
NE Persian	+1	-1	+1	+1	+1	+1	-1	-1	+1	-1	-½	-½	+1	-1	-1	+1	-½	+½	-1	-1
SA Pipil	+½	+1	+½	+1				-1	+1	-1	-1	-1		-1		0	-1		-1	-1
SA Quechua	-½	-1	-1	+1		-½	-½	0	-½	-1		-1	-1	-½	-½	-1	-1	-1	-1	-1
NA Shoshone	-1	-1	-½	+½		+½	-½	0	-½	-½		-1	-1	-1	-1	-½	-1	-1	-1	-1
NA Slave	+1	-1	-½	+1	+½	+½	-1	-1	+1	-1	+½	+½		-1			-1		-1	-1
NA Squamish		+1	+½	+½	0			+½	+1	-½	-1	+½		-1			+1		-1	-1
NE Sumerian	-½	-1	+½	+1	+½	+½	-½	+½	+1	+½		0	-1	-1	-½	+½	-1	-1	-1	-1
OC Tagalog		+1	+½	+1				-1	-1			+½	-1	-1		-1	-1	-1	-1	+½
EA Tamil	-1	-1	-1	+1	0	-1	-1	0	-1			-1	-1	+1	-½	0		+1	-1	-1
NG Tauya		-1	-½	+1	0			-1	+1	-1		-1	-1	0		-1		+½	-1	N
NE Turkish	+½	-1	-½	+1	0	-1	-1	0	-1			-½	-½	-1	+½	0	-1	+½	-1	-1
SA Tzutujil	+½	+1	+½	+1	0	-½	+½	-1	+1	-1	-½	-1	0	-1	0	+½	-1	-1	-1	N
EU Welsh	+1	+1	+1	+½	+1	+1	-1	-½	+½	-½	+1	+1	+1	+1	+1	+1	+1	+1	+1	-1
AF Wolof	-1	+½	+1	+1				-1	+1	+1	0	+1	-1	-1	-1	-1	-1	-1	+½	+1
SA Yagua	+1	+1	0	+1	0	0	-1	+½	-1			+1	-1	-1	-1	-1		-1	-1	N
NG Yimas	+1		-½	+1				+½	+1	+1		+½	-1	-1	-1	+½	-1	+½	-1	N
AF Yoruba	-½	+½	+1	+1	+1	-½	-½	-1	-1				-1	-1	-1	+1	-½	-1	-1	+1
	1	2a	2b	3	4a	4b	4c	5	6	7	8	9	10	11	12	13	14	15	16	17

Table 1b: Compressed raw scores, full sample

An initial look at these tables reveals little more than a jumble of numbers; only the blanks stand out clearly, a sign that for many languages some of the features are conceptually irrelevant. The remainder of this chapter is devoted to exploring the significance of these numbers for the Celtic/Hamito-Semitic problem.

6.2 Whole-language evaluation: summing the features

As already remarked, the scoring schema outlined in Chapter 5 was designed with a broader goal in mind than the individual features. The feature scores for a given language can be summed to provide an intuitively reasonable overall measure of degree of conformity to the ideal Celtic/Hamito-Semitic macrotype. Here the issue of oppositeness, stressed so strongly in sec. 5.1.2, comes to the fore: a language will be similar to CHS in some ways, dissimilar in others, and similarities and dissimilarities should both contribute to the overall score. The only way to attain this goal is to assign + and - scores and allow them to cancel each other out. Zeroes, quite properly, will contribute nothing to this sum: they represent neither agreement nor disagreement. Note that under this schema, two criteria must be met in order to achieve a high score: (a) a high absolute number of + scores, and (b) a strong preponderance of + over - scores. Both desiderata seem intuitively reasonable, and indeed they meet the minimal requirement for such scoring, namely that the CHS languages themselves should score high. Suspending either criterion threatens to yield counterintuitive results. Criterion (a) without (b) would allow languages with balanced high numbers of + and - scores to score high; criterion (b) without (a) would

allow a language to score high even if most of its features scored "irrelevant". We return to this issue below.

Note parenthetically that this approach to whole-language scoring is not sensitive to the particular configuration of features found in a given language; it pays no attention to the degree to which particular languages may resemble one another on a feature-by-feature basis. It does not, for example, start with Welsh (as "best" language), determine which language L most closely resembles Welsh in its feature configuration, determine which language L' best resembles L, etc. Two languages can thus attain the same overall score and yet have a very different distribution of pluses and minuses over the features. One of the most striking manifestations of this is diachronic: Ethiopic and Irish both have undergone major structural changes over a millennium of evolution, yet in neither case has the change led to any significant alteration in total score. For full discussion see sec. 6.5 below.

Any scoring schema specified in terms of a combination of multiple subfeatures must address the question of the weighting factors to be applied to the various features. In principle, this is an issue not merely for those features which were deliberately split in Chapter 5 (word order, relative-clause type), but for all of the features in the list. And it should be clear that any answer to this problem will to a considerable degree be arbitrary. I will work with two partially contradictory principles: first, that independent features should be assigned the same weight; second, that a feature cluster reflecting a richly articulated part of the language's grammar deserves to be assigned more weight than an isolated feature. The more complex and

ramified the grammatical domain, the greater weight it should have in the overall sum.

The first principle amounted to a default case: all other things being equal, apply a weighting factor of 1. The second principle came into effect for word order and relative clauses; it might well have been appropriate for feature [8] as well (article positioning in genitives), but here I chose to focus exclusively on one narrow aspect of the domain. In Chapter 5, I split word order into two features (clause word order, NP word order), and here I will retain those two features, assigning each a weight of 1.¹ Relativization is an even more complex phenomenon than word order, and hence should be assigned more weight. But the seven features and subfeatures (3,4aAbBc, 5) required to achieve a fine enough descriptive grain surely assign too much weight to the phenomenon as a whole. As already indicated, I have reduced this to the equivalent of five features by taking 4aA as a single feature and 4bB likewise; that is, the Copying/Gapping behavior of genitival relatives will count as a single averaged feature-equivalency, and likewise for adpositional relatives, as already adumbrated in the discussion of feature [4(d)] in Chapter 5. --- Possibly these procedures assign too much weight to word order and to relativization; at least the bias will apply across-the-board.

¹ It might be thought that the retention here of two distinct word order features, clearly not independent, is illegitimate in a summation assessing "degree of similarity". But a different interpretation is possible: the domain of word order as a whole is being assigned a single score, the average of the two parameters = $(F1+F2)/2$; this is then weighted double, as befits the richness of the domain. The mathematical result is the same, $F1+F2$.

The assignment of quantitative measures of whole-language resemblance is not something for which standard canons have been worked out in linguistics.² I opted for the above procedure as a simple and reasonable one, but also tried two others (to be discussed below). The results are laid out in Table 2ab, including the figures gotten by all three scoring systems. The two versions present the same information from two points of view: Table 2a is sorted on column 1, Table 2b on column 3, thus presenting the languages in two possible rank-orderings. Column 1 reflects the straightforward summation procedure just described; this constitutes the primary measure of "CHS"-ness within this study, and the reader should focus first of all on the numbers in that column, and on the rankings in Table 2a.

Key to Table 2:

- Column 1: Sum of all + and - scores
- Column 2: Number of features showing a + score
- Column 3: Sum of just those features having a + score
- Column 4: Sum of all + and - scores, counting blanks as -1
- Column 5: Number of blanks

The data distribution that emerges from Table 2a is striking.³ The

² The only attempt, or sketch of an attempt, I have seen is in the context of Balkan linguistics. Birnbaum 1970, drawing on work by Uspensky (1968) and Civ'jan, proposes an ideal model of a Balkan language (Uspensky's "étalon language"), to be taken as a standard against which actual Balkan languages can be measured quantitatively (1970:49ff., 61). The proposal, however, is only a programmatic one.

³ Tables 2a, 2b include the entire 70-language sample. Recall that six of these (Amharic, Coptic, Geez, Hebrew, Modern Arabic, Modern Irish) are excluded from the core sample; they will be touched on only marginally in the discussion below.

Whole-language scores (sorted Column 1)

	Col 1 +/-sum	Col 2 +count	Col 3 +sum	Col 4 +/-sum (I=-1)	Col 5 blank count
EU Welsh	12.0	16	15.0	12.0	0
EU Irish (Old)	10.0	15	14.0	10.0	0
EU Irish (Mod)	9.5	15	13.5	8.5	1
AF Egyptian	8.0	13	12.5	6.0	2
AF Berber	8.0	11	10.0	1.0	7
NE Arabic (Cl)	5.5	13	12.5	5.5	0
NE Hebrew	3.5	12	11.5	3.5	0
NA Chumash	2.5	8	6.5	-4.5	7
NE Arabic (Mod)	1.0	11	10.0	1.0	0
AF Geez	1.0	11	8.5	0.	1
AF Amharic	1.0	11	8.5	1.0	0
AF Hausa	1.0	11	8.0	0.	1
NG Yimas	0.5	8	6.0	-4.5	5
NA Squamish	0.5	7	5.0	-6.5	7
AF Lango	0.	9	8.5	-2.0	2
NE Akkadian	-0.5	9	8.0	-1.5	1
NE Abkhaz	-0.5	9	8.0	-0.5	0
SA Hixkaryana	-0.5	5	4.5	-7.5	7
NG Mangarayi	-0.5	5	4.0	-8.5	8
NE Persian	-1.0	9	8.5	-1.0	0
AF Wolof	-1.0	8	7.0	-4.0	3
AF Gbeya	-1.5	8	5.0	-4.5	3
AF Coptic	-2.0	9	8.0	-2.0	0
EA Gilyak	-2.0	6	4.5	-7.0	5
NA Miwok (Lake)	-2.0	4	3.5	-12.0	10
NG Hua	-2.0	4	3.0	-8.0	6
EA Lahu	-2.5	5	3.0	-9.5	7
NA Cree	-2.5	3	2.5	-11.5	9
SA Tzutujil	-3.0	7	5.0	-3.0	0
NA Eskimo	-3.0	6	4.5	-6.0	3
SA Mixtec	-3.0	5	4.0	-11.0	8
NG Dyirbal	-3.0	3	2.5	-13.0	10
AF Yoruba	-3.5	6	5.5	-6.5	3
NE Sumerian	-3.5	8	5.0	-4.5	1
NA Slave	-3.5	7	5.0	-7.5	4
SA Yagua	-3.5	5	4.5	-6.5	3
NA Kiowa	-3.5	3	2.0	-11.5	8
EU Basque	-4.0	6	4.5	-5.0	1
AF Afar	-4.0	6	4.5	-7.0	3
SA Pipil	-4.0	5	4.0	-10.0	6
NA Koasati	-4.0	5	4.0	-8.0	4
OC Fijian	-4.0	4	3.5	-10.0	6
AF Nubian	-4.0	4	3.0	-9.0	5
NA Maricopa	-4.0	3	2.5	-12.0	8
EA Chrau	-4.0	3	2.5	-13.0	9

Table 2a: Three whole-language scoring schemes (continues next page)

	Col 1 +/-sum	Col 2 +count	Col 3 +sum	Col 4 +/-sum (I=-1)	Col 5 blank count
EA Burushaski	-4.5	6	5.0	-6.5	2
AF Maasai	-4.5	5	5.0	-6.5	2
AF Mandinka	-4.5	4	4.0	-8.5	4
OC Tagalog	-4.5	5	3.5	-11.5	7
EA Chinese	-4.5	5	3.5	-10.5	6
AF Nkore-Kiga	-5.0	8	6.0	-5.0	0
AF Babungo	-5.0	6	5.0	-8.0	3
NG Tauya	-5.0	3	2.5	-11.0	6
OC Hawaiian	-6.0	6	4.5	-9.0	3
AF Hottentot	-7.0	5	3.5	-9.0	2
NE Turkish	-7.0	4	2.5	-9.0	2
NG Amele	-7.5	5	3.5	-9.5	2
EA Tamil	-7.5	3	3.0	-10.5	3
EA Japanese	-7.5	4	3.0	-10.5	3
NG Kobon	-9.0	4	3.0	-11.0	2
NE Ingush	-9.0	2	1.5	-14.0	5
EU Albanian	-9.5	5	4.0	-9.5	0
EU Hungarian	-9.5	2	1.0	-10.5	1
EU French	-10.0	5	4.5	-10.0	0
EU English	-10.5	4	2.5	-11.5	1
NA Shoshone	-11.5	2	1.0	-13.5	2
SA Quechua	-12.0	1	1.0	-14.0	2
NE Hittite	-12.0	1	1.0	-15.0	3
NE Georgian	-13.0	2	2.0	-14.0	1
EU Greek (C1)	-13.5	2	1.5	-14.5	1
Mean	-3.59		4.74		
Std Dev	5.07		2.98		

Table 2a: Three whole-language scoring schemes (sorted Column 1)

Whole-language scores (sorted Column 3)

	Col 1 +/-sum	Col 2 +count	Col 3 +sum	Col 4 +/-sum (I=-1)	Col 5 blank count
EU Welsh	12.0	16	15.0	12.0	0
EU Irish(Old)	10.0	15	14.0	10.0	0
EU Irish(Mod)	9.5	15	13.5	8.5	1
AF Egyptian	8.0	13	12.5	6.0	2
NE Arabic(C1)	5.5	13	12.5	5.5	0
NE Hebrew	3.5	12	11.5	3.5	0
AF Berber	8.0	11	10.0	1.0	7
NE Arabic(Mod)	1.0	11	10.0	1.0	0
AF Geez	1.0	11	8.5	0.	1
AF Amharic	1.0	11	8.5	1.0	0
AF Lango	0.	9	8.5	-2.0	2
NE Persian	-1.0	9	8.5	-1.0	0
AF Hausa	1.0	11	8.0	0.	1
NE Akkadian	-0.5	9	8.0	-1.5	1
NE Abkhaz	-0.5	9	8.0	-0.5	0
AF Coptic	-2.0	9	8.0	-2.0	0
AF Wolof	-1.0	8	7.0	-4.0	3
NA Chumash	2.5	8	6.5	-4.5	7
NG Yimas	0.5	8	6.0	-4.5	5
AF Nkore-Kiga	-5.0	8	6.0	-5.0	0
AF Yoruba	-3.5	6	5.5	-6.5	3
NA Squamish	0.5	7	5.0	-6.5	7
AF Gbeya	-1.5	8	5.0	-4.5	3
SA Tzutujil	-3.0	7	5.0	-3.0	0
NE Sumerian	-3.5	8	5.0	-4.5	1
NA Slave	-3.5	7	5.0	-7.5	4
EA Burushaski	-4.5	6	5.0	-6.5	2
AF Maasai	-4.5	5	5.0	-6.5	2
AF Babungo	-5.0	6	5.0	-8.0	3
SA Hixkaryana	-0.5	5	4.5	-7.5	7
EA Gilyak	-2.0	6	4.5	-7.0	5
NA Eskimo	-3.0	6	4.5	-6.0	3
SA Yagua	-3.5	5	4.5	-6.5	3
EU Basque	-4.0	6	4.5	-5.0	1
AF Afar	-4.0	6	4.5	-7.0	3
OC Hawaiian	-6.0	6	4.5	-9.0	3
EU French	-10.0	5	4.5	-10.0	0
NG Mangarayi	-0.5	5	4.0	-8.5	8
SA Mixtec	-3.0	5	4.0	-11.0	8
SA Pipil	-4.0	5	4.0	-10.0	6
NA Koasati	-4.0	5	4.0	-8.0	4
AF Mandinka	-4.5	4	4.0	-8.5	4
EU Albanian	-9.5	5	4.0	-9.5	0
NA Miwok(Lake)	-2.0	4	3.5	-12.0	10
OC Fijian	-4.0	4	3.5	-10.0	6
OC Tagalog	-4.5	5	3.5	-11.5	7

Table 2b: Three whole-language scoring schemes (continues next page)

	Col 1 +/-sum	Col 2 +count	Col 3 +sum	Col 4 +/-sum (I=-1)	Col 5 blank count
EA Chinese	-4.5	5	3.5	-10.5	6
AF Hottentot	-7.0	5	3.5	-9.0	2
NG Amele	-7.5	5	3.5	-9.5	2
NG Hua	-2.0	4	3.0	-8.0	6
EA Lahu	-2.5	5	3.0	-9.5	7
AF Nubian	-4.0	4	3.0	-9.0	5
EA Tamil	-7.5	3	3.0	-10.5	3
EA Japanese	-7.5	4	3.0	-10.5	3
NG Kobon	-9.0	4	3.0	-11.0	2
NA Cree	-2.5	3	2.5	-11.5	9
NG Dyirbal	-3.0	3	2.5	-13.0	10
NA Maricopa	-4.0	3	2.5	-12.0	8
EA Chrau	-4.0	3	2.5	-13.0	9
NG Tauya	-5.0	3	2.5	-11.0	6
NE Turkish	-7.0	4	2.5	-9.0	2
EU English	-10.5	4	2.5	-11.5	1
NA Kiowa	-3.5	3	2.0	-11.5	8
NE Georgian	-13.0	2	2.0	-14.0	1
NE Ingush	-9.0	2	1.5	-14.0	5
EU Greek (Cl)	-13.5	2	1.5	-14.5	1
EU Hungarian	-9.5	2	1.0	-10.5	1
NA Shoshone	-11.5	2	1.0	-13.5	2
SA Quechua	-12.0	1	1.0	-14.0	2
NE Hittite	-12.0	1	1.0	-15.0	3
Mean	-3.59		4.74		
Std Dev	5.07		2.98		

Table 2b: Three whole-language scoring schemes (sorted Column 3)

mean (core sample only) is -3.6, standard deviation 5.1, with scores ranging from -13.5 to +12.0. Of course, no language attains a perfect score, since the comparison is against an ideal CHS macrotype. Insular Celtic and Egyptian and Berber, to no surprise, are the best matches to the macrotype.⁴ What does come as a surprise is the dramatically high scores of these languages over against the bulk of the sample:

Welsh	OIr	Egypt	Berb	Arab	Hebr	Geez	Akk
12.0	10.0	8.0	8.0	5.5	3.5	1.0	-0.5

--- One st.dev. above mean = 1.5 (5 CHS are higher; also 1 non-CHS)

--- Two st.dev. above mean = 6.6 (4 CHS are higher, no others)

Arabic follows next, with a noticeable break; then Hebrew and the other Semitic languages. This is already noteworthy. As Morris-Jones had proposed (1900:639), Arabic is indeed a worse match than its cousin Hamito-Semitic languages. Within Semitic, note that Hebrew and especially Geez and Akkadian have scores considerably lower than Arabic; thus Arabic demonstrates the best that a Semitic language can do in the way of CHS-style syntax.

Following Arabic and Hebrew comes a language of the West Coast of North America, Chumash; then the Afroasiatic language Hausa (West Africa), matched by three Semitic languages outside the core sample, and

⁴ That the Celtic languages score better than the HS languages reflects nothing more than the fact that there are only two Celtic languages in the sample. All the CHS features by definition must occur in some Celtic and some HS language; and with only two Celtic languages to pick from, the features (and their + scores) cannot avoid piling up in Welsh or Irish or both. In HS, by contrast, the features and + scores can be parcelled out over a larger number of languages, thus yielding a lower total score for each.

followed by Yimas (New Guinea), Squamish (Pacific Northwest Coast), Lango (East Africa), and others.⁵

	<u>Chum</u>	<u>Hausa</u>	<u>Yimas</u>	<u>Squam</u>	<u>Lango</u>	<u>Akkad</u>	<u>Abkh</u>	Hixk	Mang	<u>Pers</u>	<u>Wolof</u>
Area	Calif	WAfr	NGuin	PacNW	EAfr	NEast	Cauc	SAmer	Aust	NEast	WAfr
Blanks	7	1	5	7	2	1	0	7	8	0	3
+/-sum:	2.5	1.0	0.5	0.5	0.0	-0.5	-0.5	-0.5	-0.5	-1.0	-1.0
+sum:	6.5	8.0	6.0	5.0	8.5	8.0	8.0	4.5	4.0	8.5	7.0

The fact that these, the highest-scoring non-CHS languages in the sample, come from all around the world demonstrates that Wagner's purely areal approach to the problem cannot be maintained. That Basque, the linchpin of Wagner's Eurafrikan group, scores quite poorly (-4.0) makes the identical point in a different way (see below). On the other hand, the fact that three of these "best matches" are from Africa hints at the quite significant role that areality does play in the distribution, a point which will be treated in detail in sec. 6.4 below. For the moment, the significant issue is that these "best matches" are in fact rather poor matches; compare their scores (+/- sum) with the scores of +8 to +12 characteristic of the best CHS languages. The point is clear. To judge from the sample, nothing remotely close to the full-blown CHS type occurs anywhere else in the world. Two Semitic languages (Arabic, Hebrew) take up a sort of middle ground, followed by a scattering of indifferently good matches from all over the globe.

⁵ The two languages not underlined in the table (Hixkaryana, Mangarayi) are not considered "best matches"; see discussion below.

Throughout Chapter 5, the inevitability of "judgment calls" in the scoring system was stressed repeatedly. How reliable are these total scores? Note that changing a single feature from +1 to -1 (or the reverse) will change a language's overall score by 2, which could affect its ranking considerably.⁶ For most languages, precise rank order is of little importance. For the CHS languages themselves, however, it is crucial that their position at the top of the distribution not be open to challenge. In this connection the shape of the scores' distribution curve takes on a degree of interest. For the 10-odd "best" languages, the scores fall off precipitously: these languages comprise perhaps one-sixth of the database, yet the scores drop from +12 to 0 (Lango), sometimes with large gaps between adjacent languages. For most of the remaining five-sixths, the scores cluster together far more densely, slowly dropping off (typically by half-points) to -13.5. Within this latter group, a small scoring error could indeed have great influence on a language's ranking. Within the top group of languages, however, the fact that the scores are more broadly spaced apart provides a degree of insurance against ranking error. Chumash, for example, could lose a full 2 points and yet only be ranked two lower (in the core sample), still well within the top group. And the CHS languages themselves, to speak figuratively, stand head and shoulders above the crowd --- whence our confidence in their top-level ranking.

⁶ The scores of Mixtec (-3.0) and Babungo (-5.0), for example, differ by only 2, but in ranking the languages are 19-24 apart; here a single feature switch from +1 to -1 could change the rank by some 20 rank points.

The six languages scoring between -0.5 and -1.0 might all be construed as moderately good matches to the CHS type. Here, however, we come up against an intuitive problem with the scoring schema, which emerges when we focus on Column 2 (number of + scores) or Column 3 (sum of + scores). Akkadian, Abkhaz, and Persian have positive scores for fully 9 features,⁷ and Wolof for 8. By contrast, the other two languages (Mangarayi, Hixkaryana) have a much lower number of + scores (only 5), yet nonetheless outscore Wolof and Persian. Something is amiss here. When we look at Column 5 (number of "I" scores), or at Table 1, it becomes clear what is going on: the two "low+" languages have a high number of blanks ("feature irrelevant"). This difficulty, raised as a theoretical concern in sec. 5.1.3, now reemerges in very concrete form. With so many blanks, these languages' total scores express not merely their degree of (non)conformity to the CHS type, but also (or even more so) a kind of metacategorical reluctance to even entertain the kinds of questions we want to ask. Note that several of the features which come up blank for these two languages [4c, 8, 12] are exotics, and hence show strong negative tendencies globally. By scoring blank, the "low+" languages thus manage to avoid one or several near-automatic negative scores which accrue to most other non-CHS languages. Their scores are in effect artificially inflated.

It was not clear to me how best to handle this problem. In Columns 3 and 4 of Table 2, two alternative scoring schemes are tried out as tentative responses to the issue. Column 3 represents the sum of just

⁷ This total of 9 is even higher than that found with the higher-scoring Chumash and Yimas (8) and Squamish (7).

the plus values; Table 2b re-sorts the data according to the rank-ordering resulting from this scheme. On this scoring system, Lango and Persian now become the highest-scoring non-CHS languages (even better than Chumash); Berber falls below Arabic and Hebrew; Chumash, Yimas, and Squamish drop several ranks; Nkore-Kiga rises markedly, despite its very low summation score (-5.0); Mangarayi and Hixkaryana do fall (properly) to a much lower ranking position. Column 4 embodies an approach which was suggested and rejected in sec. 5.1.3: to lump "I" (irrelevant) in with full-fledged non-CHS values, and assign it a score of -1 rather than zero. I made no actual use of data from this column.

This problem of "irrelevant" values remains an unsolved difficulty with quantitative measurements of interlanguage resemblance. If there were no blanks in the table, the straightforward +/- summation procedure would surely be the most appropriate; at the same time, it appears that a language should get credit simply for the fact of having many + scores.⁸ One might hope to improve the scoring by dividing the sum by the number of non-blank scores, thus creating a ratio; this would have the merit of systematically neutralizing the influence of blanks in computing whole-language scores. Under this scheme, however, identical scores would be assigned to a language having twelve +1's and four -1's as to one having three +1's and one -1, which is exactly what we do not want to do. Or one might instead multiply the sum by a weighting factor proportional to the number of non-blank scores; I have not done this. None of these approaches feels like the "right answer".

⁸ We cannot simply penalize a language for having many blanks; the CHS language Berber has the identical high number of blanks (7) as Hixkaryana.

What I did instead was to make an eclectic judgment appealing to both scoring schemes together. I have separated off 14 languages as top scorers:

5 CHS languages: Welsh, Irish, Egyptian, Berber, Arabic

7 best (Col 1): Chumash, Hausa, Yimas, Squamish, Lango, Akkad, Abkhaz

7 best (Col 3): Lango, Persian, Hausa, Akkad, Abkhaz, Wolof, Chumash

This yields a post facto criterion which will distinguish precisely these languages: the language must have a Column-1 score of -1.0 or better and a Column-3 score of +5.0 or better. Most of these languages have a low number of blanks; but Chumash, Squamish, and Berber have 7.

Given the central role that Basque has played throughout the history of research into the Celtic/Hamito-Semitic problem, it will be worthwhile to look closely at what the data reveal about this language. As just remarked, Basque is not a close match to the CHS type, scoring about halfway down the ranked list; its score is -4.0, slightly below the mean. Basque scores positive for 6 features (recall also sec. 2.4.2.3):

3	RC1 linker	+1	13	DO periphrastic	+1
6	Polypersonal V	+1	14	AdvCl = and + FinCl	+1/2
12	Adp periphrastic	+1/2	15	VN/Inf for FinCl	+1/2

Basque lacks a relative pronoun (3), and has a polypersonal verb (6). Additionally, it forms a progressive from a verb nominalization with the locative ending -te (12); does have a DO periphrastic (13); builds "because" clauses with a clitic eta which is homophonous with "and" (14); and has clause chaining in the sense that the conjugated Aux only

appears on the last of a chain of verbs (15). Contrary to Wagner, Basque does not really have word-initial change; to be sure, some diminutives are formed by consonant changes, but these consonants are not constrained to initial position in the word. --- Of the remaining features, Basque scores 0 on one, namely feature [5] (Special relative verb form); verbs in relative clauses do use a special form, but one found in subordination contexts of all sorts. For all other features, Basque scores negative, and the negative scores strongly outweigh the six positive scores just noted.

A final remark about languages at the very bottom of the list. The set of CHS features is intended to embody the ways in which the CHS languages differ from the bulk of Indo-European. And indeed, the old IE languages (Hittite, Classical Greek) score at the bottom of the list, with most of the modern IE languages (Albanian, French, English) doing little better; only Persian, perhaps under Arabic influence, earns a high score. These "worst" languages, it should be noted, do not form a type in the same way that the CHS languages do, even though both sets of languages occupy the extreme fringes of the distribution. The reason is that the strongly negative scores of old IE represent nothing more than an accumulation of numerous "no" values of strongly skewed, exotic privative features. This is not a legitimate basis for defining a linguistic type, for such "no" values convey only absence; they express nothing that is specially characteristic of these languages, any more than would a statement asserting that several individuals are not albino, do not come from Albania, and do not have perfect pitch.

6.3 Distribution of individual features

6.3.1 Distribution, skewing, and multivalency

As emphasized repeatedly throughout the previous chapters, one of the most important aspects of this kind of typological study is the relative numerical strength of the various values of an individual feature, that is, the shape of the distribution itself taken as a graph. Each of the feature values corresponds to a given subpopulation of the languages in the sample; how many languages are there in each such subpopulation? In the terms introduced in sec. 1.6.3, is the distribution flat (all populations roughly of the same size), skewed-inclusive (the CHS type belonging to a strong majority), or skewed-exclusive (the CHS type a small minority vis-a-vis a different strong majority)? For which features can the CHS languages be taken as "bucking a global trend"?

The raw data pertinent to this question appears in Tables 3a-3f, where each feature's distribution is presented in both a short and a long version. The short version appears on the left of the table: here Column 1 gives the number of languages having a positive score for the given feature; Column 2, true 0; Column 3, negative; Column 4, unscorable (N or I). The long version, on the right of the table, breaks these groups down further, as shown. The tables present the information from three perspectives, twice over:

		+	0	-	N/I	+1	+1/2	0	-1/2	-1	N	I
1	Conjugated Adp	26	1	26	11	17	9	1	6	20	0	11
2a	WO: clause	28	1	31	4	17	11	1	7	24	0	4
2b	WO: NP	30	4	29	1	18	12	4	19	10	0	1
3	RC1: linker	53	2	4	5	48	5	2	0	4	0	5
4a	RC1, Gen: gap/copy	20	13	8	23	13	7	13	1	7	1	22
4b	RC1, Adp: gap/copy	12	4	25	23	7	5	4	10	15	0	23
4c	RC1, Adp: move Adp	10	0	31	23	3	7	0	9	22	0	23
5	RC1: special form	21	11	31	1	6	15	11	4	27	0	1
6	Polypersonal V	34	2	28	0	33	1	2	5	23	0	0
7	Infixing/suffixing	8	2	32	22	6	2	2	9	23	0	22
8	Def Art in Gen	7	4	18	35	5	2	4	6	12	2	33
9	Nonconcord of V	13	1	37	13	7	6	1	12	25	0	13
10	VN or Inf	12	5	31	16	9	3	5	4	27	0	16
11	Pred particle	12	5	47	0	7	5	5	5	42	0	0
12	Adp periphrastic	14	1	32	17	6	8	1	7	25	0	17
13	DO periphrastic	17	10	30	7	8	9	10	1	29	0	7
14	AdvCl = and + FinCl	17	0	26	21	10	7	0	6	20	0	21
15	VN/Inf for FinCl	19	0	29	16	6	13	0	4	25	0	16
16	Word-initial change	5	0	59	0	4	1	0	4	55	0	0
17	Kin terms	13	0	40	11	6	7	0	3	37	11	0

Absolute statistics with CHS languages

Table 3a

		+	0	-	N/I	+1	+1/2	0	-1/2	-1	N	I
1	Conjugated Adp	20	1	26	11	11	9	1	6	20	0	11
2a	WO: clause	23	1	30	4	12	11	1	7	23	0	4
2b	WO: NP	24	4	29	1	12	12	4	19	10	0	1
3	RC1: linker	47	2	4	5	44	3	2	0	4	0	5
4a	RC1, Gen: gap/copy	15	13	8	22	8	7	13	1	7	1	21
4b	RC1, Adp: gap/copy	8	4	23	23	3	5	4	9	14	0	23
4c	RC1, Adp: move Adp	8	0	27	23	1	7	0	9	18	0	23
5	RC1: special form	18	10	29	1	6	12	10	3	26	0	1
6	Polypersonal V	29	2	27	0	29	0	2	4	23	0	0
7	Infixing/suffixing	6	2	28	22	4	2	2	8	20	0	22
8	Def Art in Gen	4	4	18	32	2	2	4	6	12	2	30
9	Nonconcord of V	10	1	34	13	4	6	1	12	22	0	13
10	VN or Inf	7	5	31	15	5	2	5	4	27	0	15
11	Pred particle	8	5	45	0	4	4	5	5	40	0	0
12	Adp periphrastic	11	1	30	16	3	8	1	7	23	0	16
13	DO periphrastic	15	10	27	6	6	9	10	1	26	0	6
14	AdvCl = and + FinCl	13	0	26	19	6	7	0	6	20	0	19
15	VN/Inf for FinCl	17	0	26	15	4	13	0	2	24	0	15
16	Word-initial change	2	0	56	0	1	1	0	4	52	0	0
17	Kin terms	9	0	38	11	4	5	0	3	35	11	0

Absolute statistics without CHS languages

Table 3b

		+	0	-	N/I	+1	+1/2	0	-1/2	-1	N	I
1	Conjugated Adp	40	1	40	17	26	14	1	9	31	0	17
2a	WO: clause	43	1	48	6	26	17	1	10	37	0	6
2b	WO: NP	46	6	45	1	28	18	6	29	15	0	1
3	RCl: linker	82	3	6	7	75	7	3	0	6	0	7
4a	RCl, Gen: gap/copy	31	20	12	35	20	10	20	1	10	1	34
4b	RCl, Adp: gap/copy	18	6	39	35	10	7	6	15	23	0	35
4c	RCl, Adp: move Adp	15	0	48	35	4	10	0	14	34	0	35
5	RCl: special form	32	17	48	1	9	23	17	6	42	0	1
6	Polypersonal V	53	3	43	0	51	1	3	7	35	0	0
7	Infixing/suffixing	12	3	50	34	9	3	3	14	35	0	34
8	Def Art in Gen	10	6	28	54	7	3	6	9	18	3	51
9	Nonconcord of V	20	1	57	20	10	9	1	18	39	0	20
10	VN or Inf	18	7	48	25	14	4	7	6	42	0	25
11	Pred particle	18	7	73	0	10	7	7	7	65	0	0
12	Adp periphrastic	21	1	50	26	9	12	1	10	39	0	26
13	DO periphrastic	26	15	46	10	12	14	15	1	45	0	10
14	AdvCl = and + FinCl	26	0	40	32	15	10	0	9	31	0	32
15	VN/Inf for FinCl	29	0	45	25	9	20	0	6	39	0	25
16	Word-initial change	7	0	92	0	6	1	0	6	85	0	0
17	Kin terms	20	0	62	17	9	10	0	4	57	17	0

Percent statistics with CHS languages (rounding errors)
Table 3c

		+	0	-	N/I	+1	+1/2	0	-1/2	-1	N	I
1	Conjugated Adp	34	1	44	18	18	15	1	10	34	0	18
2a	WO: clause	39	1	51	6	20	18	1	12	39	0	6
2b	WO: NP	41	6	50	1	20	20	6	32	17	0	1
3	RCl: linker	81	3	6	8	75	5	3	0	6	0	8
4a	RCl, Gen: gap/copy	25	22	13	37	13	12	22	1	12	1	36
4b	RCl, Adp: gap/copy	13	6	39	39	5	8	6	15	24	0	39
4c	RCl, Adp: move Adp	13	0	46	39	1	12	0	15	31	0	39
5	RCl: special form	31	17	50	1	10	20	17	5	44	0	1
6	Polypersonal V	50	3	46	0	50	0	3	6	39	0	0
7	Infixing/suffixing	10	3	48	37	6	3	3	13	34	0	37
8	Def Art in Gen	6	6	31	55	3	3	6	10	20	3	51
9	Nonconcord of V	17	1	58	22	6	10	1	20	37	0	22
10	VN or Inf	12	8	53	25	8	3	8	6	46	0	25
11	Pred particle	13	8	77	0	6	6	8	8	68	0	0
12	Adp periphrastic	18	1	51	27	5	13	1	12	39	0	27
13	DO periphrastic	25	17	46	10	10	15	17	1	44	0	10
14	AdvCl = and + FinCl	22	0	44	32	10	12	0	10	34	0	32
15	VN/Inf for FinCl	29	0	44	25	6	22	0	3	41	0	25
16	Word-initial change	3	0	96	0	1	1	0	6	89	0	0
17	Kin terms	15	0	65	18	6	8	0	5	60	18	0

Percent statistics without CHS languages (rounding errors)
Table 3d

		+	0	-	N/I	+1	+1/2	0	-1/2	-1	N	I
1	Conjugated Adp	49	1	49	--	32	16	1	11	37	--	--
2a	WO: clause	46	1	51	--	28	18	1	11	40	--	--
2b	WO: NP	47	6	46	--	28	19	6	30	15	--	--
3	RCl: linker	89	3	6	--	81	8	3	0	6	--	--
4a	RCl, Gen: gap/copy	48	31	19	--	31	17	31	2	17	--	--
4b	RCl, Adp: gap/copy	29	9	60	--	17	12	9	24	36	--	--
4c	RCl, Adp: move Adp	24	0	75	--	7	17	0	21	53	--	--
5	RCl: special form	33	17	49	--	9	23	17	6	42	--	--
6	Polypersonal V	53	3	43	--	51	1	3	7	35	--	--
7	Infixing/suffixing	19	4	76	--	14	4	4	21	54	--	--
8	Def Art in Gen	24	13	62	--	17	6	13	20	41	--	--
9	Nonconcord of V	25	1	72	--	13	11	1	23	49	--	--
10	VN or Inf	25	10	64	--	18	6	10	8	56	--	--
11	Pred particle	18	7	73	--	10	7	7	7	65	--	--
12	Adp periphrastic	29	2	68	--	12	17	2	14	53	--	--
13	DO periphrastic	29	17	52	--	14	15	17	1	50	--	--
14	AdvCl = and + FinCl	39	0	60	--	23	16	0	13	46	--	--
15	VN/Inf for FinCl	39	0	60	--	12	27	0	8	52	--	--
16	Word-initial change	7	0	92	--	6	1	0	6	85	--	--
17	Kin terms	24	0	75	--	11	13	0	5	69	--	--

Percent of non-N/I, with CHS languages (rounding errors)
Table 3e

		+	0	-	N/I	+1	+1/2	0	-1/2	-1	N	I
1	Conjugated Adp	42	2	55	--	23	19	2	12	42	--	--
2a	WO: clause	42	1	55	--	22	20	1	12	42	--	--
2b	WO: NP	42	7	50	--	21	21	7	33	17	--	--
3	RCl: linker	88	3	7	--	83	5	3	0	7	--	--
4a	RCl, Gen: gap/copy	41	36	22	--	22	19	36	2	19	--	--
4b	RCl, Adp: gap/copy	22	11	65	--	8	14	11	25	40	--	--
4c	RCl, Adp: move Adp	22	0	77	--	2	20	0	25	51	--	--
5	RCl: special form	31	17	50	--	10	21	17	5	45	--	--
6	Polypersonal V	50	3	46	--	50	0	3	6	39	--	--
7	Infixing/suffixing	16	5	77	--	11	5	5	22	55	--	--
8	Def Art in Gen	15	15	69	--	7	7	15	23	46	--	--
9	Nonconcord of V	22	2	75	--	8	13	2	26	48	--	--
10	VN or Inf	16	11	72	--	11	4	11	9	62	--	--
11	Pred particle	13	8	77	--	6	6	8	8	68	--	--
12	Adp periphrastic	26	2	71	--	7	19	2	16	54	--	--
13	DO periphrastic	28	19	51	--	11	17	19	1	50	--	--
14	AdvCl = and + FinCl	33	0	66	--	15	17	0	15	51	--	--
15	VN/Inf for FinCl	39	0	60	--	9	30	0	4	55	--	--
16	Word-initial change	3	0	96	--	1	1	0	6	89	--	--
17	Kin terms	19	0	80	--	8	10	0	6	74	--	--

Percent of non-N/I, without CHS languages (rounding errors)
Table 3f

Table 3a: Absolute totals, core sample

Table 3b: Absolute totals, excluding the 6 CHS languages

Table 3c: Percentage totals, core sample

Table 3d: Percentage totals, excluding the 6 CHS languages

Table 3e: Percentages ignoring N/I, core sample

Table 3f: Percentages ignoring N/I, excluding the 6 CHS languages

Tables 3ab present absolute counts of languages (a single table in lieu of 20 individual histograms); Tables 3cd, percentages calculated on the basis of all the languages; Tables 3ef, percentages calculated ignoring languages scored N or I. This last calculation is given in order to assist in computing the degree of skewing: among those languages which do have a true numerical score (Table 3ef), does the CHS type represent a minority vis-a-vis a dominant majority, and if so by how much?

Presenting the data tables in pairs (3ab; similarly 3cd, 3ef) makes it possible to view at a glance the CHS languages' contribution to the total distribution. It also provides a crude solution to the problem of low-level genetic redundancy within the sample (Celtic, Semitic, see sec. 4.4), while avoiding arbitrary decisions about which languages to suppress. Of course, the CHS languages are still overrepresented in Tables 3a,c,e. In the discussion below, accordingly, when overall language percentages are mentioned, they will usually be an approximate average of the percentages given in Tables 3ef (or 3cd). Nothing of importance is lost by this imprecision. The seeming exactness of the numbers is illusory anyway; it should be clear after the discussion in Chapter 5 how very many instances of intermediate types, competing types, hard-to-classify types, etc. have been forced (with varying

levels of confidence) into a Procrustean bed to arrive at the "exact" figures in the master data table (Table 1). Our concern, fortunately, is only to establish in broad terms what the distribution of the scores looks like for each feature; exact numbers and percentages are important only as a means to that end.

There is one major category of information which is missing from these tables: the degree to which a naturally multivalent feature has been forced into bivalency through the scoring procedures used.⁹ This issue of artificial bivalency, as emphasized repeatedly in Chapter 5, can completely alter our perception of a given feature distribution as skewed-inclusive or not. The problem in fact manifests itself in a rather subtle way, which I will address briefly here in general terms before turning to the individual features themselves. The question about skewed-inclusive distribution translates, in the present scoring schema, into a question about the number of languages attaining positive scores (+). Here the difference between +1 and +1/2 is of crucial importance --- a difference which is visible only in the long display in the tables (Table 3b). Note that for nine of the features, the non-CHS languages having positive scores are dominated by those which only score +1/2, not +1; this holds strongly for five features (4c, 5, 12, 13, 15), weakly for four others (4b, 9, 14, 17). Of course, a score of +1/2 can represent merely some limited-scope version of the selfsame phenomenon as that found in the CHS languages. However, it will be recalled that this score can also be imposed as the result of a linearization decision

⁹ Recall the reason for adopting such procedures: to be able to easily assess languages and features as to degree of resemblance to Celtic and Hamito-Semitic.

made in order to reduce a multivalent feature to a bivalent one; that is, the score $+1/2$ is often a code for some other phenomenon, not identical to the CHS one but only similar to it. In such cases, it is only a score of $+1$ which actually corresponds to the CHS phenomenon as such. For most of the nine features just mentioned, looking at the leftmost column of the short display conveys a misleading impression: many of these positive scores (those which are $+1/2$) reflect a phenomenon which is not the CHS one. In the discussion below, therefore, close attention will be paid to breaking apart the various subtypes artificially thrown together by the scoring system, and especially the distinct subtypes subsumed in the columns labeled $+1$, $+1/2$.

A final point before turning to the individual features. A heavily skewed distribution need not mean that there exists only one minority subpopulation vis-a-vis the single dominant majority. There may be several distinct minorities. What matters is the presence of a strong majority type.

6.3.2 Features

1) Conjugated adpositions

About 17% of the languages have no category "adposition" (score "I"); of the remainder, slightly under half have conjugated adpositions. No problems of linearization are involved; a score of $+1/2$ does not indicate a distinct phenomenon but rather some attenuating circumstance --- e.g. only certain adpositions may count as "conjugated", or only in certain person-number combinations, or "hedged adpositions" may be

involved, etc.

2a) Clause-level word order

Here an explicit linearization problem does come into play: SVO languages were assigned a score of $+1/2$, VSO $+1$, and indeed all the languages with $+1/2$ scores (eleven altogether) were SVO. On the other hand, scores of $-1/2$ did not normally indicate a special word-order type, but only a flexible version of SOV. About 6% of the languages had free word order (score "I"); the distribution of the remainder is roughly 25% V-first, 20% SVO, and 55% SOV. This dominance of VSO is artifactual: V-first languages, as remarked in sec. 4.4, are deliberately overrepresented in the sample, and VSO is generally considered much less common than SVO.¹⁰ Only very weak skewing (2:1) is involved.

2b) NP-level word order

Given that this feature represents an averaging of four distinct bivalent parameters (N-Adj, N-Gen, N-RC1, Adpositional polarity), the various $1/2$ scores represent nothing more than degree of approximation to the ideal extreme types (Head-Dept, Dept-Head). Thus there is no linearization problem. The two types are more or less on a par numerically, with the Dept-Head type marginally dominant. --- A score of "I" indicates free word order, described for only one language (Dyirbal).

¹⁰ For example, Hawkins's 336-language "Expanded Sample" includes 53 V-first languages vs. 109 SVO languages (1983:288).

3) Relative clause linker

As remarked, this is a multivalent feature which was reduced to bivalency by lumping. Recall that even within the CHS languages, various types of linker appear (zero; invariant; marking role in matrix clause). I will not attempt here a further breakdown of the various subtypes involved. Whatever the more detailed articulation of this feature may be, it is clear that (as regards the CHS value) the distribution is strongly skewed inclusive. The minority type, rather, is the IE-style relative pronoun (6% of the sample). --- The score "I" is given to the single language described as having no adnominal relative clauses (Hixkaryana), and to those having correlative or internally headed RCls.

4a) Genitival RCl: gapping and copying

With genitival RCls, the dominant type is "Irrelevant": a bit over a third of the languages do not relativize genitives. Of the remainder, Copying (+1) outnumbers Gap/Copy (0), which in turn outnumbers Gapping (-1): percentages are roughly 45%, 33%, 20% respectively. Scores of +1/2 do not code a distinct phenomenon, but typically reflect an unclear situation or the cooccurrence of competing strategies of different type, with +1/2 representing an average. Here mild inclusive skewing is involved, with the CHS type (Copying) weakly in the majority.

4b) Adpositional RCl: gapping and copying

Again, "Irrelevant" is the leading value: somewhat more than a third of the languages do not relativize on adpositional objects.

Scores of +1/2 cause no difficulty, representing multiple strategies, unclear situations, or "hedged" adpositions. In sharp contrast to genitival RCl's, adpositional RCl's include no examples of the Gap/Copy type; scores of zero (true 0) reflect either internally-headed/correlative RCl's, or an average of two strategies. Among the remaining languages, the dominant type would appear to be Gapping (-1) over Copying (+1), by 2.5 to 1. This skewing, however, is partly an artifact. As discussed above (sec. 5.2 [4g]), three different subtypes have been coded identically (-1): Move-Adp, pure Gapping, Relative-Pronoun. Breaking these apart, the overall percentages (ignoring Irrelevant) become roughly:¹¹

Copying 26% Move-Adp 21% Gapping 32% Rel-Pron 10% Zero 10%

The feature thus shows little strong skewing when viewed as a multivalent feature; the distribution is fairly flat, with Rel-Pronouns trailing. The CHS languages, it will be recalled, fall into two subpopulations here (Copying and Move-Adp) --- neither of which could remotely be characterized as "bucking a trend".

4c) Adpositional RCl: Move adposition

Again (recall [4b]), a bit more than one-third of the languages fail to relativize on adpositional objects. Of the remainder, nonmovement of the Adp is strongly favored, comprising about 3/4 of the non-"I" languages; this already shows a degree of skewing. In this feature, however, +1/2 usually indicates a rather different construction than

¹¹ The computation is crude; for many of the languages, several types coexist, or the phenomena are unclear, or adpositions are a marginal category.

does +1 (recall the discussion in sec. 5.2 [4f]). Only those languages where the Adp itself moves are scored +1; the score +1/2 indicates a distinct phenomenon, namely the replacement of the Adp by an adverbial element appearing beside to the verb. The full-fledged Move-Adp phenomenon seen in Berber and Old Irish (+1) recurs clearly only in Abkhaz; one other language (Nkore-Kiga) has full-fledged Move-Adp alongside other strategies, while two others (Amele, Gbeya) are unclear but seem sometimes to show this type. Four languages (Afar, Fijian, Hawaiian, Tzutujil) show the adverbial type of movement (+1/2). Despite the uncertainties, the result is a solidly skewed-exclusive distribution: the full-fledged Move-Adp phenomenon as seen in the CHS languages is a strong minority type (perhaps 11% of the non-"I" languages), over against the dominant nonmovement type (76%). The ratio is around 1:7.

5) Special relative form of verb

"Irrelevant" is only scored for the single language lacking adnominal relative clauses (Hixkaryana). In this feature, the value +1/2 has little significance: it generally indicates that a special RC1 form exists but only occurs part of the time. This score is commoner than +1; only 6 languages (none CHS) seem to make use of a special RC1 form under all circumstances. Rather, it is the score 0 which denotes a distinct type, viz. the existence of a special form used for subordinate-clause verbs in general. The overall distribution shows only very minor skewing: about 33% have a relative form, 50% do not, and the remaining 17% get a zero score. The skewing ratio is about 1:1.5.

6) Polypersonal verb

Here things appear very clear-cut: slightly over half the languages have a verb coding two or more actants, the rest do not. This feature, of course, is the result of lumping together the values of a naturally polyvalent feature. But in fact the polyvalent feature, too, has only mild skewing. Marking of two arguments on the verb is commonest (35%), while around 22% of the languages (each) mark zero, one, or three+ arguments.

7) Infixing/suffixing alternation

A bit over 1/3 of the languages score "Irrelevant" by virtue of not coding Object; not surprisingly, these all score negative on feature [6] (recall the discussion in sec. 5.2). The remaining languages show skewing: a little over three-quarters have no Object-marker alternation of any kind. In fact, however, the skewing is more extreme, as several patterns have been subsumed under "infixing/suffixing alternation". The Yimas alternation involves a prefix, not an infix; in Sumerian and Cree, the alternation is not conditioned by a preverb. Only three non-CHS languages show infixing/suffixing Object alternation of the same kind seen in Irish and Berber: Albanian, French, and Wolof. These represent about 10% of the non-"I" languages, a strongly skewed-exclusive distribution vis-a-vis the clear majority type represented by total nonalternation (75%). The ratio is 1:7.5, approximately.

8) Definite article in genitive embeddings

Some 55% of the languages scored "I" (irrelevant, reflecting lack of an article), far and away the highest "I" percentage of any of the features. Of the remainder (only 29 languages in the core sample), 65% scored negative and 20% positive. Scores of +1/2 do not reflect a different phenomenon, but either "hedged" articles (Slave) or cooccurrence of competing alternative constructions (English). The skewing ratio is about 1:3.5. This is surprisingly flat for what seems intuitively a highly marked configuration; I suspect it may be an artifact of the small sample size.

9) Nonconcord with full-NP subject

About one-fifth of the languages score "Irrelevant" by virtue of not coding subject on the verb. Of the remainder, about 23% score positive, 73% negative. This would appear to be a mildly skewed distribution of about 1:3. However, the positive scores in fact embrace a wide variety of constructions. Six of the 10 non-CHS languages which score positive have a score of +1/2, and four of these represent the "flexible" type referred to in sec. 5.2 [9f], where either N or V or both may code plurality. This seems a different phenomenon from the CHS nonconcord pattern. In Yimas (+1/2), all verbal pronominal affixes are optional. Squamish (VSO) does show the CHS pattern, but only optionally (+1/2), and variation in word order does not play a role. Four non-CHS languages score +1: Afar, Lango, Wolof, and Yagua. But unlike the CHS languages, in Wolof and Lango (SVO) and Afar (SOV) the nonconcord has nothing to do with variation in word order. In Yagua (VSO) variation in word order is crucial, but the conditioning is exactly backwards from

the CHS languages: nonconcord now occurs precisely when the full-NP subject is prenominal (SVO).

Thus a detailed look at the positive-scoring languages shows that no non-CHS language is a true match for the CHS type, in which nonconcord occurs with VSO order but not with SVO order. Rather, the scoring system has lumped together a variety of distinct types. As regards the "strong" CHS phenomenon (narrowly construed), the skewing is quite strong: the strict CHS pattern occurs in only 3 languages (all CHS themselves), amounting to perhaps 6% of the non-"I" languages --- over against a dominant majority type (48%) involving full concord and scoring -1. The ratio is thus around 1:8.¹² If, on the other hand, we construe the phenomenon broadly and count all +1 scores as full matches, the ratio becomes roughly 1:4.5.

10) Verbal noun or Infinitive

One quarter of the sample scored "Irrelevant" for this feature (1/3 are isolating, 2/3 use exclusively finite forms). Of the remainder, 68% have Infinitives, around 20% have VNs, the remainder score 0 (indeterminate, or both). (For this feature, half-scores can be lumped in with full scores; they typically indicate a VN and an Inf construction competing in the same language, or a nonfinite construction which in a minor way deviates from "canonical" VN or Inf.) The feature shows

¹² The types involving nonconcord controlled by animacy, or nonconcord only with conjoined or counted NPs (-1/2), are here kept apart from languages having full concord (-1). If these types were lumped together, the skewing would be much stronger.

moderate skewing, with ratio of the CHS type vis-a-vis the dominant majority type a bit under 1:3.5.

11) Predicative particle

This is a feature where the scores +1/2, 0, -1/2 were explicitly used to linearize various distinct phenomena; hence most (not all) instances of these scores should be split off and looked at separately (see discussion in sec. 5.2). None of the languages scores "Irrelevant". About 2/3 of the languages have a full -1 score (no predicative particle); I will count languages scoring -1/2 (multipurpose particle used for predication) as a distinct type. At the other extreme comes the CHS type, those languages having a clearly profiled predicative particle identical to an adposition (two languages scoring +1/2 also fall under this header); these comprise around 11% of the sample. Three languages scoring +1/2 have a predicative particle not identical to an adposition (5%). The five languages scoring 0 all have a particle used for verbal as well as copular predication (8%). Overall, the CHS type (narrowly construed) emerges with a rather strong degree of skewing vis-a-vis the clearly dominant type where no predicative particle occurs; the ratio is about 1:6.

12) Adpositional periphrastic

A bit over a quarter of the languages score "Irrelevant" for this feature (reflecting either lack of nonfinite forms or lack of adpositions). Of the remainder, about 70% score negative (absence of the

feature); -1/2 scores, usually indicating "hedged adpositions", will be taken here as falling into the same type as -1 scores. As for the positive scores, +1/2 signals one of various "nonstandard" versions of the construction (recall sec. 5.2): a nonadpositional particle (French, Japanese), a verbal abstract in a local case (Basque, Turkish), a locative copula (Abkhaz, Hausa, Lahu), or unusual semantics (Hixkaryana). Those languages which do have a full-fledged adpositional periphrastic include, in addition to three CHS languages, two African languages (Afar, Mandinka) and Chinese. This last group (+1) comprises about 10% of the non-"I" languages,¹³ a clearly skewed-exclusive distribution vis-a-vis the majority type (70%). The ratio is roughly 1:7.

13) "DO" periphrastic

10% of the languages score "Irrelevant" (lack of nonfinite forms). Of the remainder, half score negative (almost all -1). The score zero, attested for 10 languages (18%), indicates a distinct type ("DO" plus a nonverbal element, often a borrowed root). Of the languages scoring +1/2, four have nonstandard semantics, while two (Afar, Amele) use a verb other than "DO"; these perhaps should count as distinct constructions from the CHS prototype, perhaps not. The "pure" CHS type (+1)

¹³ Interesting figures emerge from Blansitt 1975, a crosslinguistic survey of progressive constructions. He cites perhaps 10 languages that have a prepositional periphrastic (none, except for Welsh and Irish, appearing in the present sample). Within his universe of discourse, namely 50-odd languages having some sort of special progressive form, these ten make up about 20%. In the present study, of course, the universe of discourse is quite different and surely much broader, namely languages having adpositions and verbal abstracts; thus we might well expect to come up with a lower percentage, as indeed we do (10%, given above).

occurs in about 12% of the non-"I" languages; if we include the six +1/2 languages just alluded to, this rises to about 24%. Depending how we lump things, then, the distribution will be either mildly (1:4) or very mildly (1:2) skewed.

14) Adverbial clause = "and" + finite clause

32% of the languages score "Irrelevant", by virtue of lacking a clause-level conjunction "and". Of the remainder, 63% score negative, 36% positive, none zero. The positive-scoring languages are split (+1, +1/2) on degree of semantic match to the CHS type: good semantic matches cover around 19%, poor semantic matches 16%. Depending on whether we count these as distinct types, the skewing will be either mild (1:3) or very mild (less than 1:2).

15) VN/Inf instead of finite main-clause form

One quarter of the languages score "Irrelevant", by virtue of having no verbal abstract that stands in clear opposition to a finite form. Three-fifths of the remainder score negative, two-fifths positive, an essentially flat distribution. However, a solid majority (well over two to one) of the positive scores are +1/2, representing languages which have a special clause-chaining form that cannot otherwise serve as a verbal abstract. This construction should surely count as something distinct from the CHS type, in which the form in question can be considered a real VN/Inf. The languages representing the true CHS type (+1) amount to about 10% of the non-"I" languages; the percentage for

+1/2 is a bit under 30%. Forming a ratio of true CHS (+1) to the dominant negative type, we get about 1:6, a respectable amount of skewing; the skewing stands out less, however, when set off against the "+1/2" type as third element (ratio about 1:3:6).

16) Word-initial change

No language scored "Irrelevant" or zero. This is a very rare feature, and the distribution was overwhelmingly skewed: around 95% negative, perhaps 5% positive. The only languages in the sample to show the phenomenon were Celtic (Welsh, Irish), Berber, Gilyak, and (partially) Wolof. The few scores of -1/2 indicate languages where the phenomenon can be said to exist at all only by an effort of will; these may safely be considered true negatives. The ratio is roughly 1:18, which is surely too extreme; the sample undoubtedly underrepresents the phenomenon (e.g. Mandinka, the Mande language in the database, lacks word-initial change, although the phenomenon is richly developed elsewhere in Mande).

17) Extended use of kin terms

Though no language scored "Irrelevant", about 17% scored "N" (no data); none scored zero. If we focus mechanically on the opposition between positive and negative scores, a medium skewing emerges: around 22% positive vs. 78% negative, yielding a ratio of about 1:3.5. Recall, however, that the scores +1/2 and -1/2 indicate nothing more than the degree to which the phenomenon is observed in the lexicon; and +1/2

merely signals the presence of four good examples, a very weak basis for proffering the phenomenon as a dominant characteristic of the language (as it is in Irish and Arabic). It will not be surprising that the positive scores are dominated (weakly) by +1/2; interestingly, full +1 scores occur only in the CHS languages and in 4 languages of Africa (Hausa, Lango, Wolof, Yoruba). If we look only at +1 scores (as measuring the "real thing"), the percentage drops from 22% to around 10%, yielding a skewing ratio somewhat over 1:7.5. On the other hand, the number of negative scores is surely too large (though to an unknown degree), because the absence of good lexical sources will automatically create a bias toward negative scores. In any event, there is undoubtedly strong skewing in this feature, and the phenomenon in question seems clearly a rare one.

6.3.3 Skewed-exclusive distribution of the CHS type

Does the ensemble of Celtic/Hamito-Semitic similarity features buck a global trend? The ratios of (CHS type)/(majority type), presented on a feature-by-feature basis in the previous section, provide data pointing to an answer. Table 4 is a summary statement of such ratios. Each feature is accompanied either by a skewed-exclusivity ratio or by a descriptive comment (such as "flat") if the skewing is very mild or nonexistent.¹⁴ Notations for strongly skewed-exclusive features appear

¹⁴ "Flat" means that no subpopulation shows strong numerical dominance over any of the others; the distribution need not be perfectly flat, but it should not show large spikes. "Mild exclusive" means that the CHS type indeed represents a minority over against a majority, but not by a strong margin (from 1:2 to 1:3.5).

	<u>Strong</u>	<u>Skewed-Exclusive</u>	<u>Other</u>
		<u>(Exotic)</u>	
1	Conjugated Adp		flat
2a	WO: clause		mild exclusive (?)
2b	WO: NP		flat
3	RCl: linker		strong inclusive
4a	RCl, Gen: gap/copy		flat (mild inclusive)
4b	RCl, Adp: gap/copy		flat (mild exclusive)
4c	RCl, Adp: move Adp	1:7	
5	RCl: special form		mild exclusive
6	Polypersonal V		flat
7	Infixing/suffixing	1:7.5	
8	Def Art in Gen		1:3.5
9	Nonconcord of V	1:4.5 (?) 1:8 (?)	
10	VN or Inf		1:3.5
11	Pred particle	1:6	
12	Adp periphrastic	1:7	
13	DO periphrastic	1:4 (?)	1:2 (?)
14	AdvCl = and + FinCl		1:2 (?) 1:3 (?)
15	VN/Inf for FinCl	1:6	
16	Word-initial change	1:18	
17	Kin terms	1:7.5	

Skewing within the features
Table 4

in the left-hand column, for other features in the right-hand column. As discussed in detail above, in all cases the ratio or descriptive comment does not simply emerge from a computation performed mechanically on the tabulated scores, but rather reflects the true multivalent distribution underlying the linearized scores found in the tables. In particular, only languages which are a more or less exact match to the CHS phenomenon are considered to belong to the CHS type. Features for which two figures are given reflect my indecision regarding how much leeway should be allowed in construing the notion "exact CHS match".

The striking aspect of these figures is the strong skewing found among features in the bottom half of the chart. These are all privative features which are truly "exotics" --- features that are special characteristics of a minority of languages, rather than features of intrinsic relevance to all or most languages. Such features involve "bucking a trend" by their very nature: the minority value (or several minority values) is/are set off against a dominant majority group defined by absence of the rare exotic trait.

Of the many features shared by Celtic and Hamito-Semitic, perhaps half have proven to be exotics. Seven of the features, perhaps eight, show a skewing relationship of 1:5 or stronger, one going as high as 1:18. And the behavior shown by the non-exotic features (right-hand column) is just as revealing. The mild skewing (where there is any) disfavors the CHS type in all cases but two: only features [3], [4a] show a skewed-inclusive distribution, and [4a] is almost flat. Feature [3] is very strongly skewed-inclusive; but it is the only such feature in the set. If these exotic features are independent or only weakly

correlated --- which does indeed appear to be the case (see sec 6.6 below) --- then the cumulative testimony of these skewings is dramatic indeed. Any single strongly skewed-exclusive feature involves "bucking a trend"; when two unrelated languages both share the same minority feature value, it counts as something unusual. Now we see it happening seven or eight times over (even ignoring weaker skewings). This is a far stronger kind of interlanguage resemblance than the mere fact of sharing a feature value. That such a thing should happen purely by coincidence is very unlikely.

Given these strong skewings, it might seem a straightforward matter to calculate the probability of the CHS cluster of phenomena occurring by chance in a language. To illustrate the method, assume three strongly skewed-exclusive features F1-F3, each having a 1:9 skewing ratio; then the probability of each minority value (M1-M3) occurring by chance is 1/10, and the probability of all three minority values cooccurring is 1/1000. Such a procedure would seem, at first glance, to be quite attractive with regard to the CHS problem; it would yield a probability on the order of ten million to one. And if one were to compute the probability of all these minority values cooccurring identically in two languages, the figure would be far smaller. The difficulty is that (reverting to our three-feature model) one could easily envision three different features F1'-F3', each having 10 values all equally probable --- a flat distribution. Here too the probability of any one value occurring is 1/10, and the probability of any three particular values cooccurring is again 1/1000. But the identical probability figure conveys totally different information in the two cases: in the second case, the probability of 1/1000 tells us nothing of interest, since all

possible values are equally (im)probable. Probability figures per se, then, cannot be the appropriate statistic in the present context. But the point is abundantly clear notwithstanding.

6.4 Areal patternings

Wagner's work put an areal stamp on the CHS problem, and indeed, in dealing with a problem of this nature, areal analysis is natural and all but inevitable. We have seen that Wagner's "Eurafrican" proposal lacks a crucial link: Basque has no claim to any special status, and without it the "European" half of Eurafrican loses all concrete support. On the other hand, as will shortly become evident, Wagner's intuitions about Africa turn out to be accurate and perceptive. Nowhere outside the CHS languages does the full-fledged CHS type recur; but in Africa, and notably in West Africa, a weak version of the type does appear to have a natural home.

Areal patterning emerges most clearly within the Afroasiatic superfamily itself:

Language:	Berb	Egyp	Arab	Hebr	Hausa	Geez	Akkad	Afar
Subgroup:	Berb	Egyp	Sem	Sem	Chad	Sem	Sem	Cush
	8.0	8.0	5.5	3.5	1.0	1.0	-0.5	-4.0

By a broad margin the scores are highest on the Mediterranean rim, falling away in every other direction: to the east (Akkadian), southeast (Geez and especially Afar), and southwest (Hausa). Note the low score of the Semitic languages Akkadian and Geez (the furthest from the Mediterranean), and the extremely low score of the Cushitic language Afar,

falling even below the sample mean of -3.6. The label "Mediterranean Hamito-Semitic", accordingly, exactly fits the distribution of the full-blown CHS type.

The surprisingly high score of the non-CHS language Hausa provides a first indication of the areal prominence of Africa as a whole, and particularly of West Africa. In Table 5a, the scores for individual languages have been sorted out by area, and average overall scores for areas computed. The results are simple and clear. The overall score for Africa (left column) is highest of all the areas listed (-1.6); even when the scores of the African CHS languages (Berber, Egyptian) are removed from the sample, the score remains high (-3.3). Relatively high-scoring languages, with scores around 0 to -2, occur in various parts of Africa; Lango, a Nilotic language of East Africa, is among the best. But a particularly high-scoring cluster exists in West Africa. Looking only at the "non-CHS" languages, the 6 West African languages in the sample (Hausa plus 5 Niger-Congo languages, viz. Wolof, Mandinka, Yoruba, Babungu, and Gbeya) do rather better than the rest of the continent (Cushitic, Eastern Sudanic, East African Bantu, Hottentot); the average West African score is -2.4.¹⁵

Indeed, several of the exotic CHS features are at home in Africa, in particular West Africa. This can be seen in part from Table 5b, which shows the percentage which each global area contributes to the total tally of + scores for each feature (the CHS languages are split

¹⁵ Africa north of Zaire (non-CHS) had an average score of -3.0; while the lowest-scoring language on the continent was Hottentot (-7.0) in the far south. Clearly the scores fall off fairly steadily on a north-south cline.

Averages of areal statistics
(sorted by +/- sum)

Area	+/-sum	+count	+sum	I-count
Africa				
Core	-1.6	7.4	6.3	2.9
West Africa	-2.4	7.2	5.8	2.8
No CHS	-3.3	6.7	5.4	2.6
North America				
	-3.1	4.8	3.7	6.2
New Guinea/Austr				
	-3.8	4.6	3.5	5.6
South/Centr Amer				
	-4.3	4.7	3.8	4.3
East Eurasia				
	-4.6	4.6	3.5	5.0
Oceania				
	-4.8	5.0	3.8	5.3
Near East				
Core	-4.6	6.3	5.4	1.4
No CHS	-6.6	5.0	4.1	1.7
Europe				
Core	-4.4	6.9	5.9	0.5
No CHS	-9.5	4.0	3.0	0.7

Legend: Core = Core sample
 No CHS = Core sample minus CHS languages

Table 5a: Areal averages

	Percentages of + scores, by areas									CHS+AF
	CHS	AF	NE	EU	EA	NG	OC	NA	SA	
1	0.23	0.15	0.15	0.04	0.08	0.04	0.	0.12	0.19	0.38
2a	0.18	0.29	0.	0.11	0.07	0.	0.11	0.11	0.14	0.46
2b	0.20	0.27	0.07	0.13	0.03	0.03	0.10	0.07	0.10	0.47
3	0.11	0.21	0.09	0.02	0.13	0.13	0.06	0.15	0.09	0.32
4a	0.25	0.35	0.10	0.	0.15	0.05	0.	0.10	0.	0.60
4b	0.33	0.33	0.17	0.	0.	0.	0.	0.17	0.	0.67
4c	0.20	0.30	0.10	0.	0.	0.10	0.20	0.	0.10	0.50
5	0.14	0.24	0.14	0.	0.05	0.10	0.05	0.24	0.05	0.38
6	0.15	0.15	0.12	0.09	0.03	0.15	0.	0.24	0.09	0.29
7	0.25	0.13	0.13	0.25	0.	0.13	0.	0.13	0.	0.38
8	0.43	0.14	0.14	0.14	0.	0.	0.	0.14	0.	0.57
9	0.23	0.23	0.	0.	0.08	0.08	0.08	0.23	0.08	0.46
10	0.42	0.08	0.25	0.	0.	0.08	0.	0.08	0.08	0.50
11	0.33	0.25	0.08	0.	0.08	0.08	0.	0.08	0.08	0.58
12	0.21	0.21	0.14	0.14	0.21	0.	0.	0.	0.07	0.43
13	0.12	0.18	0.12	0.12	0.06	0.24	0.	0.12	0.06	0.29
14	0.24	0.06	0.	0.18	0.12	0.	0.06	0.29	0.06	0.29
15	0.11	0.16	0.11	0.05	0.21	0.32	0.	0.05	0.	0.26
16	0.60	0.20	0.	0.	0.20	0.	0.	0.	0.	0.80
17	0.31	0.38	0.	0.08	0.15	0.	0.08	0.	0.	0.69
Tot	0.21	0.22	0.09	0.06	0.09	0.09	0.04	0.13	0.08	0.42

	Percentages of +1 scores, by areas									CHS+AF
	CHS	AF	NE	EU	EA	NG	OC	NA	SA	
1	0.35	0.18	0.18	0.	0.06	0.06	0.	0.06	0.12	0.53
2a	0.29	0.06	0.	0.06	0.	0.	0.18	0.18	0.24	0.35
2b	0.33	0.33	0.06	0.11	0.06	0.	0.06	0.	0.06	0.67
3	0.08	0.21	0.10	0.02	0.15	0.15	0.06	0.13	0.10	0.29
4a	0.38	0.38	0.08	0.	0.08	0.	0.	0.08	0.	0.77
4b	0.57	0.29	0.14	0.	0.	0.	0.	0.	0.	0.86
4c	0.67	0.	0.33	0.	0.	0.	0.	0.	0.	0.67
5	0.	0.17	0.17	0.	0.17	0.17	0.	0.33	0.	0.17
6	0.12	0.15	0.12	0.09	0.03	0.15	0.	0.24	0.09	0.27
7	0.33	0.17	0.	0.33	0.	0.17	0.	0.	0.	0.50
8	0.60	0.20	0.20	0.	0.	0.	0.	0.	0.	0.80
9	0.43	0.43	0.	0.	0.	0.	0.	0.	0.14	0.86
10	0.44	0.	0.33	0.	0.	0.11	0.	0.	0.11	0.44
11	0.43	0.29	0.	0.	0.14	0.	0.	0.	0.14	0.71
12	0.50	0.33	0.	0.	0.17	0.	0.	0.	0.	0.83
13	0.25	0.25	0.13	0.25	0.	0.13	0.	0.	0.	0.50
14	0.40	0.	0.	0.10	0.	0.	0.10	0.40	0.	0.40
15	0.33	0.33	0.	0.	0.33	0.	0.	0.	0.	0.67
16	0.75	0.	0.	0.	0.25	0.	0.	0.	0.	0.75
17	0.33	0.67	0.	0.	0.	0.	0.	0.	0.	1.00
Tot	0.28	0.21	0.09	0.05	0.07	0.07	0.03	0.11	0.08	0.50

Table 5b: Areal percentages (+)

off as a separate "area").¹⁶ A quick perusal of the bottom subtable will show that Africa contributes a disproportionate percentage of +1 values (full-fledged matches to CHS) for numerous exotic features.¹⁷ Adpositional periphrasis, for example, is common in Africa, though perhaps not particularly so in the sample; see Blansitt 1975, Heine et al. 1991:153, and Heine & Reh 1984.¹⁸ The metaphorical use of kin terms is another African areal feature, present in Hausa, Lango, Wolof, and Yoruba, and also in the Kwa language Ewe [Johanna Nichols, p.c.]. The sample underrepresents the importance of word-initial change in West Africa: it is richly developed in southwestern Mande (Welmers 1973:130ff.), though nonexistent in Mandinka, and much more highly profiled in Fula (not in the sample) than in its relative Wolof; also the existence of word-final change in the Nilotic language Nuer is noteworthy (Lieber 1987:78ff.). In East Africa, it should be noted, another exotic feature is quite common: the "DO" periphrastic, with the areal twist (characteristic of Cushitic [Leslau 1945:72]) that the Aux

¹⁶ Reading across the first row (Feature [1]), the figures means that, of all languages scoring + for this feature, 23% are CHS languages, 15% African (excluding CHS), etc. The "Total" row gives a similar percentage for the aggregate of all instances of a language scoring +. The second table presents the same information, but looking only at languages which are a full-fledged match for the CHS phenomenon (+1).

¹⁷ This is not just a result of the disproportionately large number of African languages in the sample. Even when the areas are grouped into macrozones (see secs. 4.4 and 6.1.1), all roughly balanced numerically, Africa still shows the same predominance, though to a lesser degree.

¹⁸ Blansitt describes a prepositional periphrastic in the African languages Margi(?), Mbum, Mambar, Diola Fogany, Umbundu, and Ngambay-Moundou (1975:9-10, 11-12, 13-14, 17, 24, 27). Heine et al. state that constructions of the type they term PP-periphrasis "have developed into progressive and similar aspects in over one hundred African languages" (1991:153) --- though this type, defined as "Subj Aux Prep NP" (1984:115), is structurally less constrained than the type "Subj Cop Prep VN/Inf" used in this study. For "localism" in African languages, see also Blok 1948.

verb is not "do" but "say".

At the opposite extreme is the situation in Europe. Its average continental score on Table 5a is low, but not extremely so (-4.4); three other areas score marginally lower. However, if the two Celtic languages are removed from the sample, the score for Europe plummets to -9.5, by far the lowest for the areas examined. The implication is clear: Irish and Welsh are much stranger in a European context than are Egyptian and Berber in an African one.¹⁹ The Near East presents a somewhat paradoxical situation reminiscent, in an attenuated way, of that seen in Europe. The paradox inheres in the presence of extremely high-scoring and extremely low-scoring languages in the same area, even ignoring the CHS languages (Semitic). The Northwest Caucasian language Abkhaz scores well (-0.5), the other Caucasian stocks very poorly; Persian does well (-1.0), Sumerian moderately well (-3.5), while the languages of Anatolia score low. The typological "schizophrenia" long associated with the Caucasus may thus be part of a broader geographical phenomenon. On the other hand, the high score of Persian may reflect long contact with Arabic. Notwithstanding Persian and Abkhaz, however, the area as a whole --- unlike Africa --- scores quite low.

Looking back to the top end of the distribution, we see that the area with the second-highest score is North America (-3.1); indeed, this score surpasses Africa without the CHS languages. And the average would be higher still (-2.2) were it not for the gap between bottom-scoring

¹⁹ Note that this European context is not exclusively Indo-European, but also includes Basque and Hungarian. Conversely, the only (non-Celtic) IE language that scores at all well falls outside Europe (Persian).

Shoshone (-11.5) and the second-lowest score (-4.0). It might seem odd that here, halfway around the world, the CHS type should score so high. Yet an examination of the last column in the chart, the count of "I" (Irrelevant) scores, reveals a fundamental difference. North American languages have the highest "I" score of all areas examined, with an average of slightly over 6 per language: they typically lack such categories as definite articles, adpositions, and nonfinite verb forms (i.e., verbs devoid of person-number coding), and seldom allow genitival and adpositional relativization. These categories are structural prerequisites for several exotic features; their absence means that the typical North American language avoids six likely -1 scores for these features, thereby seriously inflating its overall score. This is a defect in the scoring system (recall the discussion in sec. 6.2). Recognizing this flaw enables us to apply an interpretive correction to the raw numbers, and see the North American patterning for what it is: not so much a categorial agreement with the CHS languages, as rather a categorial indifference to many of the structural questions that define the type.²⁰

In summary, the areal analysis reveals several basic points. The categories which are definitional to the CHS type are well installed chiefly in Africa, Europe, and the Near East.²¹ However, though sharing

²⁰ That the languages score moderately high rather than moderately low is in part a reflection of the predominance of Head-marking in North America: most of the languages have a polypersonal verb, and those that have the category "adposition" usually have conjugated adpositions. Special RC1 forms of the verb are also common, which might be construed as a kind of head marking (in that the verb, as head of the RC1, codes the function of the clause as a whole); further, the conjunction "and" often has adverbial nuances.

²¹ The somewhat higher "I" score of Africa may be attributed to the

the appropriate categorial prerequisites, these regions reveal marked differences in how the CHS features are realized, with Africa set off sharply against the other two regions. The CHS type, in its weak form, finds a home in Africa; the Near East is typologically bifurcated, but generally low-scoring; in Europe the CHS type is radically out of place. These considerations harmonize well with a prehistoric account involving some link specifically between Insular Celtic and Africa. By contrast, to see the rise of the CHS type in Britain as a purely Europe-internal matter would run directly counter to the typological and areal facts.

The areal data also argue against the plausibility of Wagner's sprachgeographical account. With Basque out of the picture, the alleged Eurafrikan "area" finds concrete support only in Insular Celtic and Hamito-Semitic, strongly reducing its evidential value. The only factor clearly favoring such a scenario is the rather high score of Abkhaz, coupled with the striking match in blood type between the territories of Abkhaz and Insular Celtic (recall sec. 1.5.3). It is tempting to take Abkhaz as an attenuated modern-day witness to an ancient CHS-like area of broad scope. But Abkhaz is geographically very remote from the "Eurafrikan" region as postulated by Wagner, and any areal explanation resting purely on Celtic, Abkhaz, and Mediterranean Hamito-Semitic is geographically tenuous on the face of it. We will return to sprachgeographical explanation at the end of the next section.

absence of definite articles and clause-level "and" conjunctions in many of the languages.

6.5 Diachrony

Almost all the CHS languages have more recent descendants --- as remarked in sec. 4.4, four such languages were deliberately included in the "supplementary" sample. Ample data, therefore, is available for considering the CHS problem in diachronic perspective. Do the languages change their degree of CHS-ness over time? If so, how?

The supplementary sample includes descendant languages for 4 of the 8 CHS languages: Old Irish (Modern Irish), Egyptian (Coptic), Classical Arabic (Modern Arabic), and Geez (Amharic).²² Comparative diachronic data for these 4 language pairs is shown in Table 6. The other 4 CHS languages do not lend themselves to a similar diachronic analysis, for a variety of reasons. Middle and Modern Welsh are almost identical as regards the CHS features. Berber is attested only as a modern language; Akkadian, only as an ancient language, which died millennia ago with no apparent descendants. Biblical Hebrew does have a modern version, contemporary Israeli Hebrew. But the unique status of Modern Hebrew as a consciously and deliberately "resurrected" language implies an inherent discontinuity in its evolution; the language is not simply the descendant of its parent language in the sense that (say) Modern Irish is the descendant of Old Irish. Indeed, Modern Hebrew syntax in many ways reflects the "Standard Average European" syntax of the languages spoken by the Zionist immigrants. Accordingly, Modern Hebrew will not be further considered here.

²² Write-ups for Amharic and Coptic are included in Appendix 2. For Modern Irish the discussion is based on Christian Brothers 1980, and for Modern Cairene Arabic on Gary & Gamal-Eldin 1982.

	1	2a	2b	3	4a	4b	4c	5	6	7	8	9	10	11	12	13	14	15	16	17
Arabic (Cl)	+1	+1	+1	+1	+1	+1	-1	-1	+1	-1	+1	+1	+1	+½	-1	-1	+1	-1	-1	+1
Hebrew	+1	+1	+1	+1	+1	+1	-1	-1	+1	-1	+1	+½	-1	-1	-1	-1	+1	+1	-1	+1
Geez	+1	+1	+½	+1	+1	+1	+½	-1	+1	-1		-½	+½	-1	-1	-1	+½	-1	-1	+½
Akkadian	+1	-1	+1	+1	+1	+1	-1	0	+1	-1		-1	+½	-1	-1	-1	+1	-½	-1	+½
Egyptian	+1	+1	+1	+1	+1	+1	-1	+½	-½	-1		+1	+1	+1	+1	+1		+1	-1	-1
Berber	+1	+1	+1	+1		-1	+1	+½	+1	+1		-1		+1					+1	+½
Irish (Old)	+1	+1	+1	+½	+1	-½	+1	+½	+1	+1	+1	-1	+1	-1	+1	-1	+1	-½	+1	+1
Welsh	+1	+1	+1	+½	+1	+1	-1	-½	+½	-½	+1	+1	+1	+1	+1	+1	+1	+1	+1	-1

	1	2a	2b	3	4a	4b	4c	5	6	7	8	9	10	11	12	13	14	15	16	17
Irish (Old)	+1	+1	+1	+½	+1	-½	+1	+½	+1	+1	+1	-1	+1	-1	+1	-1	+1	-½	+1	+1
Irish (Mod)	"	"	"	"	"	+1	-1	+1	-1		"	+1	+½	+½	"	"	"	-1	"	"
Welsh	"	"	"	"	"	"	"	-½	+½	-½	"	"	+1	+1	"	+1	"	+1	"	-1
Arabic (Cl)	+1	+1	+1	+1	+1	+1	-1	-1	+1	-1	+1	+1	+1	+½	-1	-1	+1	-1	-1	+1
Arabic (Mod)	"	+½	"	"	"	"	"	"	"	"	+½	-1	"	-1	"	"	"	"	"	"
Geez	+1	+1	+½	+1	+1	+1	+½	-1	+1	-1		-½	+½	-1	-1	-1	+½	-1	-1	+½
Amharic	"	-1	-1	"	"	+½	"	0	"	-½	+1	-1	-1	"	+1	+½	"	"	"	"
Egyptian	+1	+1	+1	+1	+1	+1	-1	+½	-½	-1		+1	+1	+1	+1	+1		+1	-1	-1
Coptic	"	+½	"	"	"	"	"	-1	+1	"	-1	+½	-1	"	-1	0	-1	-1	"	"

1 2a 2b 3 4a 4b 4c 5 6 7 8 9 10 11 12 13 14 15 16 17

Scores:

Irish (Old)	10.0	Arabic (Cl)	5.5	Egyptian	8.0	Geez	1.0
Irish (Mod)	9.5	Arabic (Mod)	1.0	Coptic	-2.0	Amharic	1.0

Table 6: Diachronic Data

The four diachronic "case studies" differ considerably among themselves. Arabic is by far the most straightforward. The language loses two CHS features entirely, [9] (Nonconcord) and [11] (Predicative Particle). Two others change from +1 to +1/2: the word order shifts from VSO to SVO [2a], and a new analytical genitive using an "of" word arises (alongside the old Construct formation), allowing both nouns to have the article [8]. And that is all. Four relatively minor alterations work together, all in the same direction, to lower the score of Arabic from 5.5 to 1.0.

Arabic represents a case of noncataclysmic linguistic evolution. What change there is all works in the same "antiextremist" direction, diluting or eliminating several hallmarks of the CHS type; the effect on the language's overall structural cut is minimal, amounting entirely to the reduction of marked syntagms. The case of Coptic is far more extreme, and differs from the Arabic example both quantitatively and qualitatively. Here the score does not merely drop but plunges: +8.0 to -2.0. And the evolutionary path which realized the change was radical. Coptic represents in effect the polysynthetic re-formation of an earlier stage of analytical syntax (see Hodge 1970). Just as Romance developed a new synthetic future from a Latin analytical construction ("Verb + GO"), so too Coptic vis-a-vis Pharaonic Egyptian, only even more so. The intricate Coptic verb is the compressed reflex of earlier Egyptian periphrastic constructions --- significantly, constructions which are definitional for the CHS type. Egyptian thus saw much of its old CHS-type syntax swallowed up in the verb; and the new Coptic syntax which evolved was not at all like the old. Conjugated prepositions do survive [1], as does N-Modifier order [2b] and the thoroughgoing use of Copying

in relativization [4]. Little else is the same. Clause-level word order switches from VSO to a variety of VSO and SVO options, including incorporation of full-NP subjects into the verb [2a]. There is (on my reanalysis; see Appendix 2) no special RCl form of the verb [5]. The complexity of the verb increases markedly, with two actants now routinely coded [6]; this is the only respect in which Coptic is more CHS-like than Egyptian. Coptic develops an article, whose positioning in genitive constructions (with an explicit "of" particle) does not conform to the special CHS pattern [8]. The verb now shows full concord with its subject, except when the Subj is incorporated [9]. The verbal abstract behaves like an Infinitive, not a VN [10]. The old prepositional periphrastic has been absorbed into the Coptic verb, and the construction has not regenerated [12]. A DO-periphrastic (DO + X) does survive, but the uninflected element X is not exclusively or even primarily deverbal [13]. Coptic develops an "and" conjunction, which does not have the characteristic CHS adverbial usage [14]; the language cannot use its Infinitive for finite predication [15]. Ten features have changed altogether, leading to a radical switch in type. Yet in spite of the precipitous drop in score, the resultant score of -2.0 is still a bit higher than the overall mean for Africa as a whole.

Ethiopic Semitic (Amharic) underwent major changes as well, though not as extreme as Coptic. Here the workings of the Cushitic substratum are very much in evidence, notably in word order: the language has moved from VSO to strongly OV [2a, 2b], though prepositions do survive (alongside new postpositions). Prepositional relative clauses (as also in Geez) can involve movement of the Prep to the Verb (a feature also seen in the Cushitic language Afar); the straightforward copying strategy is

less prominent than in Geez [4b, 4c]. The language develops something like a special relative form [5]. It has evolved an article which does conform to the CHS genitive patterning [8], but as part of a more general rule whereby the article appears exclusively on the prenominal modifier in any Modifier-N syntagm. Full concord is the rule (Geez allowed nonconcord) [9]. The verbal abstract is an infinitive, not a VN [10]. Amharic acquires an adpositional periphrastic (cf. Afar) [12]. And, in striking conformity to a quirk of the Cushitic substratum, it acquires a "DO-periphrastic" which features the verb "say", not "do" (thus e.g. in Afar; see explicitly Armbruster 1960:30, 192).

The most curious aspect of the ten changes undergone by Amharic vis-a-vis Geez (the same number as for Coptic) is that they all cancel out. Half of the features decline in score (total -6.0) and half gain (+6.0), including polarity reversals (+ to - or vice versa) in five features, yielding a total identical to that of Geez (1.0). The detailed profile of the changes is even more interesting and highly instructive, for here the workings of a substratal language are patent. About half the change --- a loss of 3.5, a gain of 3.5 --- is due to substratal influence [features 2a, 2b, 13, probably 12]. Several of the CHS features explicitly cross "deep" genetic lines here. Feature [13] is especially noteworthy: the fact that the Aux is "say", an unusual and marked trait,²³ is a serendipitous diagnostic arguing against independent parallel development.

²³ Though not unparalleled globally; "say+X" is the normal periphrastic construction in Yuman (see write-up for Maricopa in Appendix 2).

The one remaining case, that of Irish, is undoubtedly the most puzzling. In gross quantitative terms, Irish is like Ethiopic: 9 features change, 5 involving polarity reversals, yet the total is barely affected (Old Irish: 10.0; Modern Irish: 9.5). In Modern Irish, Prep RCIs take on the copying strategy characteristic of most CHS languages and lose the "Move Adp" strategy dominant in Old Irish [4b, 4c] --- thus at least for the standard Irish dialect described in the grammar book (Christian Brothers 1980), though in Scots Gaelic the "Move Adp" construction survives (McCone 1985:96).²⁴ Regarding further instances of lowered scores, Modern Irish does not code the object on the verb [6], hence cannot show infixing/suffixing alternation [7]; the VN has a strongly developed secondary use as an Infinitive (nongenitival rection) [10]; and the VN cannot be used as an independent predication [15]. Of particular interest, however, are instances of raised scores vis-a-vis Old Irish. In addition to the Prep RC1 behavior just noted, the special relative form of the verb is less restricted than in Old Irish [5] (1980:144-45); verb nonconcord has become the rule [9]; and a distinctive predicative-particle construction with "in", unknown in Old Irish, has arisen [11]:

He is in his doctor = "He is a doctor".

The last two constructions are puzzling because they postdate Old Irish and hence cannot plausibly be attributed to a substratum (recall sec. 3.1). Adstratal influence from English would not yield such

²⁴ McCone proposes that this choice between two types of Prep RC1 may have been a structural isogloss separating northern and southern Irish even in Old Irish times, with "Old Irish" as it has come down to us (employing the "Move Adp" strategy) being based on the northern, Ulster dialect (1985:97; also Ahlqvist 1988:28).

constructions. Adstratal influence from Welsh (or other Brythonic) on Irish is a tempting explanatory approach: the development from Old to Modern Irish involves three changes that could be construed as "bringing Irish into line with Welsh" (Prep RCIs, Nonconcord, probably Pred Particle [4b, 4c, 9, 11]) --- though a comparative feature profile of Modern Irish vs. Welsh shows almost as many differences between the two languages as does a profile of Old Irish vs. Welsh (Table 6). Intimate Welsh-Irish contact on a large scale, however, seems unlikely in historical times; Brythonic was the language of Britain, not Ireland. Why then did these features [9, 11] arise? The possibility immediately suggests itself that they are perhaps natural typological concomitants of the CHS type --- as if Old Irish, already a strongly CHS-type language, was under internal structural pressure to become even more fully "CHS" in character. For feature [9] this may be plausible; as we will see below (sec. 6.6.1), this feature does show a correlation (albeit not unimpeachable) with verb-firstness and Head-marking, hallmarks of the CHS type. But no such correlation appears for feature [11]. We return to this issue in sec. 7.4.

Taken together, these four individual case studies hint at intriguing generalizations. First, however, the four must be augmented with a fifth: that of Insular Celtic itself vis-a-vis older Indo-European, where an enormous type shift occurred, a swing of some +20 points. From these five case studies several conclusions emerge. First of all, the type is very mutable. It can wax (Insular Celtic) and wane (Coptic) drastically; it can also remain stable in overall score even while shuffling the deployment of the individual features (Amharic, Irish). Interestingly, the (non)mutability does not correlate with the presence

of substratal activity. The drastic type change seen in Insular Celtic vis-a-vis IE surely did involve a substratum (whatever its identity), but the change in Coptic probably did not; on the other hand, the overall nonchange in Amharic (individual features change, overall score does not) was accompanied by considerable substratal re-formation, whereas the change from Old to Modern Irish, displaying much the same pattern, presumably involved no substratal action.

Second, CHS features (including exotic features) demonstrably can cross genetic lines, as seen by Cushitic influence on Ethiopic in features [2a], [2b], [13], and probably [12].

Third and perhaps most significantly, if we consider Mediterranean Hamito-Semitic as a whole, it is noteworthy that the three case studies presented all involve either decrease or parity in total score over long time periods. No cases of increase occur in this part of the world. Combined with the African areal conclusions laid out in the previous section, this yields the tentative projection that the ancestral languages of northern and western Africa may well have been even better exemplars of the CHS type than are their modern-day descendants: the observed trend, as exemplified in Egyptian and Ethiopic (and Arabic), is for African languages to get less CHS-like over time. Accordingly, the languages of northern and western Africa millennia ago --- the time when a putative migration to Britain might have taken place --- stand a fair chance of having been just as CHS-like as they are now, and perhaps more so. On this view, Berber can be taken as an extraordinary modern survival of a once-commoner type, and ancient Egyptian, which in synchronic perspective looks typologically extreme vis-a-vis the modern languages

of West Africa, may have been much less of a structural oddity in its own time. That this is guesswork will be obvious, but the guesses are plausible; there appears to be nothing in Africa pointing in the opposite direction. Indeed, the vast geographical sweep of the (weak) CHS type in modern Africa, from Wolof in Gambia to Lango in Uganda, itself points to CHS-ness as something very old in this part of the world.

This diachronic projection has implications for Wagner's sprachgeographical account as well. Wagner's scenario posits, over most of Western Europe, a radical type change away from an earlier CHS-like type, a change allegedly resulting from typological obliteration under a wave of Indo-European invaders. But would such a wave have eradicated almost every trace of the earlier type? Unless the Indo-Europeans killed or kept aloof from the local population --- which seems unlikely --- one might have expected substratal and other contact effects to set in everywhere that the invaders penetrated. Some weak degree of CHS behavior might have been expected to persist in Europe, rather than a sweeping continent-wide type inversion. Such weak persistence is just what we do see in the northern half of Africa, which geographically and linguistically provides a strikingly good synchronic analogue to Wagner's hypothetical ancient Eurafrikan area: both are large regions, continental in scale, encompassing many CHS-type languages of diverse genetic affiliation. Plausibly, the behavior of the (weak) CHS type as actually observed in Africa should yield insight into the probable behavior of the type in Wagner's alleged Eurafrikan area. In Africa the type is surely old and well-installed, and it does cross major genetic lines: Niger-Congo (Wolof), Afroasiatic (Hausa), Nilotic (Lango). Clearly, whatever the prehistoric movements of peoples in Africa may

have been, they did not prevent the (weak) CHS type from prevailing in many languages over vast distances and over very long periods, in the process penetrating several major linguistic stocks. By contrast, Wagner's scenario suggests that, from a similar starting point, a totally different evolutionary scenario worked itself out in Western Europe: the CHS type, originally well-installed just as in Africa, did not show broad staying power but instead succumbed almost everywhere. --- Of course, there is one salient difference: ancient Africa, so far as is known, did not ever undergo an invasion by speakers of an Indo-European-type language, so alien to the CHS type. Thus the African parallel cannot disprove Wagner's scenario. But it does make it somewhat less likely by demonstrating that, in the one attested parallel situation, the (weak) CHS type proved to be resilient and genetically assimilable all over the region in question.

6.6 Feature independence and correlation

The investigation has left one crucial point unresolved: to what extent can we say that the 20 features are independent variables? Answering this question will require a method of correlating features with one another, and additionally with the major linguistic types known to typologists: word order and head-dependent marking. These issues will be addressed in turn in the next two sections, followed by a detailed examination of three particularly interesting feature pairs.

6.6.1 Correlations between features

Establishing the degree of correlation between features would appear, at first glance, to be a fairly straightforward procedure: develop an algorithm and simply apply it to all feature pairs. Two factors, however, complicated the process considerably.

The first difficulty, the problem of large linguistic areas (Dryer 1989), is intrinsic to any attempt to reach linguistically meaningful global correlations. In establishing correlations or implicational relations between features in a language sample, linguists typically (though not necessarily) aim at uncovering general truths about how human language works. To contribute usefully to this goal, such generalizations must apply to human language itself, not to any particular geographical or genetic group of languages. However, as Dryer points out, correlations which hold globally may in fact arise largely because of area-specific patternings. A single continent --- Africa, for instance --- may show a strong correlation between two features which, elsewhere in the world, vary more or less randomly. A correlation algorithm applied globally to these features will show them to be correlated; but such a correlation, though it may illuminate our linguistic knowledge of Africa, will tell us nothing about human language in general.

In the context of the present study, questions about whether two features should count as independent are certainly intended as questions about human language in general. Of course some degree of global correlation is highly likely to exist, if only because of the contribution of the six CHS languages with their highly distinctive shared typological

profile. We have also seen that a weak form of the CHS pattern appears to have a natural home in Africa, and hence we might expect African languages to show some interfeature correlation, too. The real issue is whether such correlations can be substantiated all over the world and hence can be ascribed to human language as a whole. It is with this question in mind that I grouped the languages of the sample into the four numerically balanced macrozones (plus CHS) introduced in sec. 4.4:

CHS (6) EU/NE (13) AF (12) EA/NG/OC (17) NA/SA (16).

And I will approach the question of feature-pair correlations with a particular "null hypothesis" in mind: that correlations may indeed be observed, but that most of the contribution to such correlations should stem from the CHS languages and from Africa. Any correlation based largely on these languages will thus be excluded from consideration. Conversely, in order to be considered unimpeachably valid, a correlation ought to hold in all 5 macrozones, or at least not have any macrozone showing a reversal of the overall trend.

The second problem has not, to my knowledge, been addressed in the typological literature, and is specific to correlations involving privative features.²⁵ An example will demonstrate very clearly the difficulty. Consider the features "coding tense-of-possession on nouns" and "having Object-first word order", two highly exotic privative features which conceptually have nothing to do with each other. We would not

²⁵ As remarked in sec. 4.2, the present study differs from most typological investigations in its particular focus on exotics rather than on features of global linguistic relevance. Such globally relevant features tend strongly to be equipollent (word order oppositions, head-dependent marking, degree of synthesis, etc.); hence the problem of correlating privative features does not arise.

expect these features to be correlated. Yet in fact languages will overwhelmingly have the identical value "no" for both features, thereby threatening to "swamp" the correlation with meaningless negative scores. The really interesting fact --- that one language, Hixkaryana, in fact comes up "yes" for both features --- runs the risk of getting buried.

Two distinct problems are intertwined here, one more serious than the other. The first has to do with correlations involving any two strongly skewed features, whether privative or not. Such features can be compared to multiple throws of two heavily weighted dice. The outcome we usually observe is the same single value over and over, yet the question of whether the dice are correlated really depends on the minority values, the ones which seldom show up. This difficulty is relatively minor: the standard product-moment correlation coefficient r , for example, turns out not to be "swamped" by a huge preponderance of one particular value. The second problem, however, is more subtle, and stems from the specifically privative nature of the exotic features in question. For features of the kind used in this study (discrete; having only a few values; basically + vs. -), normal correlation algorithms will judge two features to be well correlated if they preferentially assume the same polarity (++, --) and only seldom show inverse polarity (+-, -+). Here ++ and -- pairings are treated on a par: both represent equally good matches. For two equipollent features, this seems entirely reasonable. For two privative features, however, these two pairings are conceptually not on a par at all. What is of interest is the ++ pairings and only the ++ pairings, with all other pairings (+-, -+, --) lumped together indifferently as "other"; the -- pairings have no claim to special status, and "same polarity" does not seem a useful notion.

This is because the negative pole of a privative feature is by definition conceptually meaningless; it conveys no positive information, merely the absence of information (recall the discussion in sec.

5.1.1).²⁶ When a correlation is swamped by negative scores (--) of this kind, most of the correlation will be contributed by numbers that do not represent anything at all in the real world. This is different in kind from a similar "swamping" phenomenon with an equipollent skewed feature, where the negative numbers do mean something.

With equipollent features, as remarked, correlations should properly consider both positive and negative values. Here we will use an algorithm comparing the fraction of same-polarity scores (++, --) with the fraction of inverse-polarity scores (+-, -+); two features will be considered well correlated if most languages show the same polarity and few show inverse polarity. When the correlation involves a privative feature and an equipollent feature, however, the negative values of the privative feature are taken to be conceptually meaningless; what we are interested in, rather, is only the positive values, and how well they agree in polarity with the equipollent feature. Here, therefore, we will consider only languages where the privative feature scores positive, focusing on the relative sizes of the populations showing ++ scores as opposed to +- scores; languages where the privative feature scores negative (-+, --) will be ignored. Finally, the correlation of two privative features poses particular problems. For one thing, very

²⁶ To say of a person in Norway that he is Ethiopian conveys information; to say of such a person that he is not Ethiopian conveys nothing (Gricean implicature aside) that would not equally be conveyed by silence.

few languages will typically come up + for both features, a fact which threatens the significance of any statement that could be made; moreover, it is unclear what the count of ++ matches should be set into opposition with. In fact I will not set it into opposition with anything, but only look at the raw counts and at the areal distribution of these ++ scores.²⁷ --- NB: For all correlations of all kinds, only languages where both features score either + or - will be considered; languages where one (or both) of the features scores 0, N, or blank will not be counted.

We thus have three quite different correlation procedures for the three different types of feature pairings: equipollent-equipollent, privative-equipollent, privative-privative. These will be treated in turn. Recall that the equipollent features are 1, 2a, 2b, 3, 4a, 4b, 6, 10.

Correlations between two equipollent features are those which are most likely to be interesting from the viewpoint of general linguistics. As indicated above, for such feature pairs a computer count was made of the number of languages for which the two features either both scored + or both -. The results are shown in Table 7a. Reading across the table, the first 5 columns give data for each of the 5 macrozones;

²⁷ The statistical problems involved in treating this data set appear to be quite complex, and the ad hoc nature of the methods used here will be apparent. Taken in conjunction with the areal criterion referred to above, however, these methods should yield a reasonable degree of insight into the issue of feature interdependencies.

		Correlations of paired equipollent features							
F1	F2	CHS	AF	NE/EU	EA/NG/OC	NA/SA	ALL	NO.CHS&AF	
2a	2b	++/--	5	10	9	12	14	50	35 Y
		+/-/+	1	0	3	1	0	5	4
		pct ++/--	0.83	1.00	0.75	0.92	1.00	0.91	0.90
		pct +1+1,-1-1						0.32	0.29
		Pct all ++/-- (CHS):	0.10					Pct all ++/-- (CHS&AF):	0.30
1	10	++/--	5	7	8	5	4	29	17 Y
		+/-/+	0	2	3	2	2	9	7
		pct ++/--	1.00	0.78	0.73	0.71	0.67	0.76	0.71
		pct +1+1,-1-1						0.69	0.71
		Pct all ++/-- (CHS):	0.17					Pct all ++/-- (CHS&AF):	0.41
4a	4b	++/--	4	4	8	1	1	18	10 R,F
		+/-/+	1	2	0	3	0	6	3
		pct ++/--	0.80	0.67	1.00	0.25	1.00	0.75	0.77
		pct +1+1,-1-1						0.67	0.60
		Pct all ++/-- (CHS):	0.22					Pct all ++/-- (CHS&AF):	0.44
1	4a	++/--	5	4	7	2	1	19	10 R,F
		+/-/+	0	3	1	3	0	7	4
		pct ++/--	1.00	0.57	0.88	0.40	1.00	0.73	0.71
		pct +1+1,-1-1						0.84	0.80
		Pct all ++/-- (CHS):	0.26					Pct all ++/-- (CHS&AF):	0.47
4b	10	++/--	4	3	7	5	1	20	13 F
		+/-/+	1	3	3	0	1	8	4
		pct ++/--	0.80	0.50	0.70	1.00	0.50	0.71	0.76
		pct +1+1,-1-1						0.55	0.46
		Pct all ++/-- (CHS):	0.20					Pct all ++/-- (CHS&AF):	0.35
1	6	++/--	5	7	5	10	8	35	23 R
		+/-/+	1	3	7	2	2	15	11
		pct ++/--	0.83	0.70	0.42	0.83	0.80	0.70	0.68
		pct +1+1,-1-1						0.66	0.65
		Pct all ++/-- (CHS):	0.14					Pct all ++/-- (CHS&AF):	0.34
4a	6	++/--	4	4	5	3	2	18	10 F
		+/-/+	1	3	3	2	0	9	5
		pct ++/--	0.80	0.57	0.63	0.60	1.00	0.67	0.67
		pct +1+1,-1-1						0.61	0.50
		Pct all ++/-- (CHS):	0.22					Pct all ++/-- (CHS&AF):	0.44
1	4b	++/--	4	3	8	7	2	24	17 R
		+/-/+	2	3	4	0	3	12	7
		pct ++/--	0.67	0.50	0.67	1.00	0.40	0.67	0.71
		pct +1+1,-1-1						0.54	0.47
		Pct all ++/-- (CHS):	0.17					Pct all ++/-- (CHS&AF):	0.29

Table 7a (continues next page)

F1	F2	CHS	AF	NE/EU	EA/NG/OC	NA/SA	ALL	NO.CHS&AF	
6	10	++/--	4	5	7	7	5	28	19 Y
		+-/-+	1	4	4	5	1	15	10
		pct ++/--	0.80	0.56	0.64	0.58	0.83	0.65	0.66
		pct +1+1,-1-1						0.64	0.63
		Pct all ++/-- (CHS):	0.14					Pct all ++/-- (CHS&AF):	0.32
4b	6	++/--	3	4	6	6	3	22	15 Y
		+-/-+	3	3	5	1	2	14	8
		pct ++/--	0.50	0.57	0.55	0.86	0.60	0.61	0.65
		pct +1+1,-1-1						0.55	0.53
		Pct all ++/-- (CHS):	0.14					Pct all ++/-- (CHS&AF):	0.32

Legend: Y = "yes" (the correlation is unexceptionable)

R = "reversal" (one area shows a reversal)

F = "few" (less than half the sample contributes to the correlation)

Table 7a: Correlations of equipollent features

column 6 totals the data from all five columns, and column 7 totals the data excluding the CHS and African languages. Reading vertically, the first row gives raw totals for same-polarity scores (++/--); the second row, for opposite-polarity scores (+/-,+/-); if we let T represent the total of these, the third row gives the percent P of same-polarity languages divided by T; the fourth row shows the percentage of same-polarity scores for which both features contribute non-half scores (+1+1, -1-1); and the fifth row shows what percent of the same-polarity scores is contributed by the CHS languages, and by the CHS languages plus the African languages. The languages are sorted according to the same-polarity percentage P; a value of P=50% indicates no correlation.

Numerous feature pairs were excluded from Table 7a, for three reasons:

- (a) If $P < 60\%$; such pairs are too weakly correlated to count.
- (b) If more than half the languages contributing same-polarity scores are CHS or African; such pairs will be taken as confirming the null hypothesis that correlations stem from CHS or African connections.
- (c) If two or more areal zones showed a polarity reversal; a correlation counterexemplified in two zones is rejected.

This left 10 pairings (below) which might plausibly be considered correlated. Two contingencies could weaken the correlation, however:

- (a) If one areal zone showed a polarity reversal; the correlation is counterexemplified in this zone.
- (b) If less than half the sample contributed to the correlation.

The former contingency is coded "R" (reversal) on the table; the latter, "F" (few); and the notation "Y" (yes) is used for a completely unexceptionable correlation. The 10 correlated pairs are summarized compactly here; the percentages P are given twice, first for the entire core sample, then for the sample excluding Africa and the CHS languages:

[2a] Clause-level WO	[2b] NP-level WO	91% 90% Y
[1] Conjug Adp	[10] VN/Inf	76% 71% Y
[4a] Gen RC1: Gap/Copy	[4b] Adp RC1: Gap/Copy	75% 77% R,F
[1] Conjug Adp	[4a] Gen RC1: Gap/Copy	73% 71% R,F
[4b] Adp RC1: Gap/Copy	[10] VN/Inf	71% 76% F
[1] Conjug Adp	[6] Polypersonal V	70% 68% R
[4a] Gen RC1: Gap/Copy	[6] Polypersonal V	67% 67% F
[1] Conjug Adp	[4b] Adp RC1: Gap/Copy	67% 71% R
[6] Polypersonal V	[10] VN/Inf	65% 66% Y
[4b] Adp RC1: Gap/Copy	[6] Polypersonal V	61% 65% Y

By far the strongest correlation is precisely the one we would have expected even in advance of this study (2a,2b): the two word-order parameters score very high, thus setting the standard for what it means to be well correlated under this measure. Only one other pairing seems solidly established (1,10): languages with conjugated adpositions tend quite strongly to have verbal nouns, and vice versa. One might expect the two Head-marking features in the sample (1,6) to show some correlation, and they do; the macrozone EU/NE constitutes an exception (indeed, both EU and NE separately show a reversal), but if the CHS languages are added back into their respective geographic zones the reversal vanishes. Since we have a link between (1,6) and (1,10), we would expect one

between (6,10); indeed one exists, and the correlation is weak but unexceptionable. The remaining correlations involve the RC1 Gap/Copy features. All such correlations are imperfect in some way: most are marked F or R, one scores quite low. We can say, in a general way, that the Copying strategy for relativization tends to be preferred with Head-marking languages (features 1,4a,4b,6 all correlate with one another) and tends to correlate with verbal nouns (4b,10). It is striking what kind of correlation is not observed: keeping to the 60% cut-off point, and adhering to the criteria laid out above, there are no correlations interrelating word order and any other equipollent feature.²⁸ Nor do any of the correlations involve feature [3] (RC1 Linker), since that feature scores positive for almost all the languages. What we do observe is a correlational network of varying strength linking verbal nouns, Head-marking, and the Copying strategy for relativization. The connection between the latter two suggests interesting directions for future theoretical research. If the Copying type indeed tends to prefer a bound pronominal copy (Head-marked), it may be because such pronouns are structurally less "obtrusive" than independent pronouns. Functionally, resumptive pronouns are strongly backgrounded elements, and a bound (as opposed to a free) pronoun would formally mirror this functional backgrounding.

²⁸ Almost all such feature-pairs have a correlation score between 50% and 60%.

We turn now to correlations between a privative feature (PF) and an equipollent feature (EF). Since only positive scores of PF are meaningful, we restrict our view to instances where PF scores +, and investigate whether EF tends to score positive or negative under this restriction. The results are laid out in Table 7b, whose format is similar to that of Table 7a. The columns are identical. The first row gives raw ++ counts, the second gives raw +- counts (NB: the privative feature comes first in these polarity pairings), the third gives raw counts for cases where the privative feature scores minus (-+, --). The fourth row gives the percentage P of ++ scores divided by all scores for which PF scores positive (++, +-). The fifth row shows the percentage of ++ matches for which both features contribute non-half scores (+1+1). Finally, the sixth row shows what percentage of ++ scores is contributed by the CHS languages, and by the CHS languages plus the African languages. The languages are sorted according to the ++ percentage P.

A great many feature pairs were excluded from Table 7b, for almost the same three reasons given for Table 7a. Again a cut-off of $P < 60\%$ was imposed, and again instances of double polarity-reversal were excluded; this time, however, I allowed the CHS and African languages to comprise as much as 60% of the total ++ score.²⁹ A further across-the-board exclusion was any correlation involving feature [3]; since feature [3] almost always came up positive, such correlations were not interesting. This left eleven pairings which might be taken as correlated. As

²⁹ The justification for doing this is that the total number of languages involved in a typical correlation was much lower than in a typical correlation on Table 7a; hence it seemed reasonable to allow a bit more leeway to compensate for the lowered significance.

		Correlations of privative and equipollent features (in that order)						
F1	F2	CHS	AF	NE/EU	EA/NG/OC	NA/SA	ALL	NO.CHS&AF
7	6	++	2	1	3	1	8	5 Y
		+-	0	0	0	0	0	0
		-+/--	4	5	5	4	12	30
		pct ++(of +x)	1.00	1.00	1.00	1.00	1.00	1.00
		pct +l+l					0.75	0.60
		Pct all ++ (CHS):	0.25				Pct all ++ (CHS&AF):	0.38
7	2b	++	2	1	3	0	6	3 R
		+-	0	0	0	1	1	1
		-+/--	4	6	6	3	12	31
		pct ++(of +x)	1.00	1.00	1.00	0.00	0.86	0.75
		pct +l+l					0.83	0.67
		Pct all ++ (CHS):	0.33				Pct all ++ (CHS&AF):	0.50
7	2a	++	2	1	2	0	1	6 3 Y
		+-	0	0	1	0	0	1
		-+/--	4	6	5	3	12	30
		pct ++(of +x)	1.00	1.00	0.67	0.00	1.00	0.86
		pct +l+l					0.50	0.33
		Pct all ++ (CHS):	0.33				Pct all ++ (CHS&AF):	0.50
9	1	++	3	1	0	2	2	8 4 R
		+-	0	2	0	0	0	2
		-+/--	3	4	12	5	7	31
		pct ++(of +x)	1.00	0.33	0.00	1.00	1.00	0.80
		pct +l+l					0.63	0.25
		Pct all ++ (CHS):	0.38				Pct all ++ (CHS&AF):	0.50
4c	2b	++	2	2	0	2	1	7 3 C,R
		+-	0	1	1	0	0	2
		-+/--	4	6	12	4	4	30
		pct ++(of +x)	1.00	0.67	0.00	1.00	1.00	0.78
		pct +l+l					0.29	0.00
		Pct all ++ (CHS):	0.29				Pct all ++ (CHS&AF):	0.57
9	2a	++	3	2	0	1	3	9 4 C
		+-	0	1	0	1	1	3
		-+/--	3	4	11	7	10	35
		pct ++(of +x)	1.00	0.67	0.00	0.50	0.75	0.67
		pct +l+l					0.44	0.25
		Pct all ++ (CHS):	0.33				Pct all ++ (CHS&AF):	0.56
15	4a	++	2	1	1	2	0	6 3 Y
		+-	0	1	0	1	0	2
		-+/--	3	5	6	1	0	15
		pct ++(of +x)	1.00	0.50	1.00	0.67	0.00	0.75
		pct +l+l					0.33	0.00
		Pct all ++ (CHS):	0.33				Pct all ++ (CHS&AF):	0.50

Table 7b (continues next page)

F1	F2	CHS	AF	NE/EU	EA/NG/OC	NA/SA	ALL	NO.CHS&AF	
13	4a	++	2	1	2	1	0	6	3 Y
		+-	0	0	1	1	0	2	2
		-+/--	3	7	4	1	0	15	5
		pct ++(of +x)	1.00	1.00	0.67	0.50	0.00	0.75	0.60
		pct +l+l						0.67	0.33
		Pct all ++ (CHS):	0.33					Pct all ++ (CHS&AF):	0.50
4c	2a	++	2	2	0	2	1	7	3 C,R
		+-	0	1	1	1	0	3	2
		-+/--	4	5	11	4	5	29	20
		pct ++(of +x)	1.00	0.67	0.00	0.67	1.00	0.70	0.60
		pct +l+l						0.29	0.00
		Pct all ++ (CHS):	0.29					Pct all ++ (CHS&AF):	0.57
5	1	++	3	3	1	2	2	11	5 C,R
		+-	0	2	2	1	0	5	3
		-+/--	2	5	8	8	4	27	20
		pct ++(of +x)	1.00	0.60	0.33	0.67	1.00	0.69	0.63
		pct +l+l						0.09	0.20
		Pct all ++ (CHS):	0.27					Pct all ++ (CHS&AF):	0.55
9	6	++	2	2	0	1	3	8	4 R
		+-	1	1	0	2	1	5	3
		-+/--	3	3	11	8	10	35	29
		pct ++(of +x)	0.67	0.67	0.00	0.33	0.75	0.62	0.57
		pct +l+l						0.38	0.00
		Pct all ++ (CHS):	0.25					Pct all ++ (CHS&AF):	0.50

Legend: Y = "yes" (the correlation is unexceptionable)

R = "reversal" (one area shows a reversal)

C = "CHS" (CHS languages contribute 51-60% of ++ scores)

Table 7b: Privative-equipollent correlations

with Table 7a, there were again numerous qualifications on the legitimacy of the correlations. The codes "Y" (yes) and "R" (reversal) mean the same as with Table 7a; additionally, the code "C" will indicate cases where the CHS languages contribute 51-60% of the ++ scores. The last column in the mini-table below gives the total number of ++,+- combinations attested for the given feature pair excluding the CHS and African languages, to provide an indication of how very few languages actually partake in these "correlations". Here are the eleven feature pairs; again, the two percentages P are for the entire core sample, and then for the sample minus Africa and CHS:

[7] Infix/suffix	[6] Polypersonal V	100%	100%	Y	5
[7] Infix/suffix	[2b] NP-level WO	86%	75%	R	4
[7] Infix/suffix	[2a] Clause-level WO	86%	75%	Y	4
[9] Verb Nonconcord	[1] Conjug Adp	80%	100%	R	4
[4c] Adp RC1: Move Adp	[2b] NP-level WO	78%	75%	C,R	4
[9] Verb Nonconcord	[2a] Clause-level WO	75%	67%	C	6
[15] VN/Inf for FinCl	[4a] Gen RC1: Gap/Copy	75%	75%	Y	4
[13] DO Periphrastic	[4a] Gen RC1: Gap/Copy	75%	60%	Y	5
[4c] Adp RC1: Move Adp	[2a] Clause-level WO	70%	60%	C,R	5
[5] Special RC1 Verb	[1] Conjug Adp	69%	63%	C,R	8
[9] Verb Nonconcord	[6] Polypersonal V	62%	57%	R	7

With so few languages involved, little can be said. First of all, a very large number of correlations were excluded by virtue of being dominated by CHS and African languages: of the 51 feature-pairs whose correlation fraction was greater than the cut-off point of 60%, fully 33 had over half their ++ values contributed by CHS or African languages

--- almost two-thirds of the relevant languages (33/51). Only 9 feature-pairs had less than half their ++ values from CHS and African languages; for the remaining 9 pairs, the fraction ("F") was exactly a half. This clearly confirms the null hypothesis adopted herein, whereby much of the observed overall correlation is posited to be contributed by Africa and the CHS languages. By contrast, it is noteworthy that the equipollent-equipollent correlations were not similarly dominated by CHS+Africa. --- The correlations shown in Table 7b, of course, are those for which $F < 60\%$, hence those that should at least have a chance of attaining some global validity. Most of these correlations are flawed in some way, being marked either "C" or "R". Interestingly, the best correlation (7,6) is essentially a tautology: all languages having an infixing/suffixing alternation must code Obj on the verb, and almost all Obj-coding languages are polypersonal. The infixing/suffixing alternation also seems to correlate (with one areal flaw) with VO macrotype. Verb nonconcord shows some (flawed) correlation with Head-marking (9,1 and 9,6) and with VO word order (9,2a). Beyond these brief comments I have nothing to add to the table itself. There is, however, another respect in which almost all these correlations are weak. For all feature-pairs except the first two in the table above, the ++ values (non-CHS, non-Africa) include at most one full-fledged +1+1 pairing; usually one or the other (or both) is +1/2.³⁰ Thus these already imperfect correlations tend to involve phenomena which are only reminiscent of, not solid matches for, the CHS phenomena.

³⁰ This is reminiscent of the situation already seen in Table 5b for the features taken individually: overall and for most of the exotic features, about half the +1 values are contributed by CHS languages and Africa, leaving relatively few +1 values elsewhere in the world.

Correlations between two privative features are beset by a special problem: only the + values of both features have any meaning, so that the only interesting case is ++. It was not at all clear to me what this ++ type could or should be set in opposition to; fortunately, it turned out hardly to matter. The program I wrote for this kind of correlation simply did areal counts of the ++ cases. The results are displayed in Table 7c, where the data are sorted by total absolute ++ count (Column 6), arbitrarily cutting off at 4. As with the privative-equipollent type, very few languages took part in such privative-privative correlations --- a maximum of 5 non-CHS non-African languages. Again the data very strongly confirmed the null hypothesis whereby the correlation is dominated by CHS and African languages. As before, if (for a given feature-pair) we let F = fraction of ++ values coming from CHS or African languages, then of the 66 possible feature-pairs, we have

41 pairs: $F > 50\%$ 13 pairs: $F = 50\%$ 10 pairs: $F < 50\%$

(2 pairs had no ++ values at all). The 41 cases where $F > 50\%$ were excluded from Table 7c; only 23 pairs were left whose ++ values were not dominated by CHS and Africa. In all but two of these, however (5-14, 4c-8), one or both of the + values was +1/2 rather than +1. Thus almost none of the 23 "areally eligible" ++ pairings involved the CHS phenomena in full-fledged form. Little purpose would be served by more sophisticated analysis: clearly almost none of the privative features show any real affinity for one another outside the CHS languages and Africa.

The results of this crude attempt at correlation can be summed up briefly. The equipollent features show a strong correlation of the two

		"Correlations" of paired privative features							
F1	F2	CHS	AF	NE/EU	EA/NG/OC	NA/SA	ALL	NO.CHS&AF	
13	15	++	2	1	2	3	0	8	5
		+/-+/-	3	9	9	6	6	33	21
		pct ++	0.40	0.10	0.18	0.33	0.00	0.20	0.19
		pct +1+1						0.38	0.00
		Pct all ++ (CHS):	0.25					Pct all ++ (CHS&AF):	0.38
5	9	++	1	1	0	2	3	7	5
		+/-+/-	4	6	10	6	6	32	22
		pct ++	0.20	0.14	0.00	0.25	0.33	0.18	0.19
		pct +1+1						0.00	0.00
		Pct all ++ (CHS):	0.14					Pct all ++ (CHS&AF):	0.29
5	15	++	1	1	0	3	1	6	4
		+/-+/-	3	9	10	7	2	31	19
		pct ++	0.25	0.10	0.00	0.30	0.33	0.16	0.17
		pct +1+1						0.17	0.00
		Pct all ++ (CHS):	0.17					Pct all ++ (CHS&AF):	0.33
12	15	++	2	1	2	1	0	6	3
		+/-+/-	3	10	10	7	6	36	23
		pct ++	0.40	0.09	0.17	0.13	0.00	0.14	0.12
		pct +1+1						0.50	0.00
		Pct all ++ (CHS):	0.33					Pct all ++ (CHS&AF):	0.50
5	14	++	1	0	0	1	3	5	4
		+/-+/-	2	7	10	6	5	30	21
		pct ++	0.33	0.00	0.00	0.14	0.38	0.14	0.16
		pct +1+1						0.20	0.25
		Pct all ++ (CHS):	0.20					Pct all ++ (CHS&AF):	0.20
5	13	++	1	1	1	1	1	5	3
		+/-+/-	3	9	9	10	4	35	23
		pct ++	0.25	0.10	0.10	0.09	0.20	0.13	0.12
		pct +1+1						0.00	0.00
		Pct all ++ (CHS):	0.20					Pct all ++ (CHS&AF):	0.40
13	14	++	1	0	1	1	2	5	4
		+/-+/-	3	7	10	4	4	28	18
		pct ++	0.25	0.00	0.09	0.20	0.33	0.15	0.18
		pct +1+1						0.20	0.00
		Pct all ++ (CHS):	0.20					Pct all ++ (CHS&AF):	0.20
9	15	++	2	0	0	2	0	4	2
		+/-+/-	3	7	11	8	7	36	26
		pct ++	0.40	0.00	0.00	0.20	0.00	0.10	0.07
		pct +1+1						0.50	0.00
		Pct all ++ (CHS):	0.50					Pct all ++ (CHS&AF):	0.50

Table 7c (continues next page)

F1	F2		CHS	AF	NE/EU	EA/NG/OC	NA/SA	ALL	NO.CHS&AF
9	14	++	2	0	0	0	2	4	2
		+/-+/-	2	5	11	4	8	30	23
		pct ++	0.50	0.00	0.00	0.00	0.20	0.12	0.08
		pct +1+1						0.50	0.00
		Pct all ++ (CHS):	0.50					Pct all ++ (CHS&AF):	0.50
5	7	++	2	0	1	1	0	4	2
		+/-+/-	3	6	7	3	8	27	18
		pct ++	0.40	0.00	0.13	0.25	0.00	0.13	0.10
		pct +1+1						0.00	0.00
		Pct all ++ (CHS):	0.50					Pct all ++ (CHS&AF):	0.50
12	17	++	1	1	0	2	0	4	2
		+/-+/-	4	9	12	7	4	36	23
		pct ++	0.20	0.10	0.00	0.22	0.00	0.10	0.08
		pct +1+1						0.25	0.00
		Pct all ++ (CHS):	0.25					Pct all ++ (CHS&AF):	0.50
12	14	++	2	0	1	1	0	4	2
		+/-+/-	2	8	11	5	3	29	19
		pct ++	0.50	0.00	0.08	0.17	0.00	0.12	0.10
		pct +1+1						0.50	0.00
		Pct all ++ (CHS):	0.50					Pct all ++ (CHS&AF):	0.50

Table 7c: "Correlations" of privative features (++)

word order parameters, and a separate system of more weakly intercorrelated features comprising Head-marking, verbal nouns, and relativization by Copying. Outside of the equipollent features, however, any attempt at correlation runs foul of an extreme paucity of data. The null hypothesis --- that most of the observed correlation can be attributed to the CHS languages and to Africa --- is strongly borne out for privative-equipollent and privative-privative pairings. Further, these latter types, and especially the privative-privative type, are dominated by 1/2 scores; hence whatever correlation is observed does not really involve the "true" CHS phenomena.

As suggested above, correlations involving a privative feature are conceptually vexed, and the method I have used is surely open to criticism. Yet the defects in the method are probably less damaging than might at first appear. As remarked, the key role in any correlation involving privative features must be played by the ++ pairings. For such pairings we have shown, first, that most of the + scores are not +1 but +1/2; second, that a very large fraction of such ++ pairings is contributed by the CHS and African languages; third, that the number of languages having such a ++ pairing is quite small; and finally, for privative-equipollent correlations, that numerous correlations are counterexemplified by some particular areal macrozone. Crucially, these considerations are logically prior to the actual process of correlation: they do not have to do with the computation of correlation coefficients, nor with the rank ordering of feature pairs as well or poorly correlated, but with the inherent evidential value of the feature pairs themselves no matter how they are ranked. They will therefore apply identically no matter what correlation process is applied to the features, and

will have the identical effect: to minimize the meaningfulness of most correlations involving privative features.

Our examination of interfeature correlations provides new confirmation that the exotic features are indeed exotic: they may sporadically recur outside the CHS languages and Africa, but only minimally do they ever occur together in the same language. The importance of the exotic features in delineating the CHS type now emerges in sharp focus. The CHS macrotype is comprised of both equipollent and privative features, but the equipollent features clearly show a degree of mutual interdependence. The exotic features, apparently, do not. What really gives the type its specialness, therefore, is the multiplicity of exotic features, among which no meaningful correlation seems evident.

6.6.2 Correlations involving types

The other major kind of question regarding independence of features is whether there exists any correlation between the major linguistic types known to typologists --- word order and head-dependent marking --- and a language's ranking for goodness-of-fit to the CHS type. Features [2a,b] indicate a language's word order type, while [1], [6] (Conjugated Adposition, Polypersonal Verb) convey information as to head-dependent marking type;³¹ hence the question can be recast as one of correlation between these four type-defining features and the overall CHS ranking. Information relevant to this and related questions appears in Table 8, which presents the raw data set (Table 1b) in a format showing only +

³¹ A third Head-Dependent parameter should be taken into account, namely marking of pronominal possessor on the possessed noun; but this feature is not in the database.

		1	2a	2b	3	4a	4b	4c	5	6	7	8	9	10	11	12	13	14	15	16	17
EU Welsh	12.0	+	+	+	+	+	+	-	-	+	-	+	+	+	+	+	+	+	+	+	-
EU Irish(Old)	10.0	+	+	+	+	+	-	+	+	+	+	+	+	+	-	+	-	+	-	+	+
EU Irish(Mod)	9.5	+	+	+	+	+	+	-	+	-	-	+	+	+	+	+	-	+	-	+	+
AF Egyptian	8.0	+	+	+	+	+	+	-	+	-	-	+	+	+	+	+	+	+	-	-	-
AF Berber	8.0	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+
NE Arabic(Cl)	5.5	+	+	+	+	+	+	-	-	+	-	+	+	+	+	+	+	+	-	-	+
NE Hebrew	3.5	+	+	+	+	+	+	-	-	+	-	+	+	+	+	+	+	+	+	+	+
NA Chumash	2.5	+	+	+	+	+	+	-	+	+	-	-	+	0	-	-	-	+	-	-	-
NE Arabic(Mod)	1.0	+	+	+	+	+	+	-	-	+	-	+	-	+	-	-	-	+	-	-	+
AF Hausa	1.0	+	+	+	+	+	0	-	+	+	-	+	-	+	-	+	-	-	-	-	+
AF Geez	1.0	+	+	+	+	+	+	+	+	+	-	-	+	-	+	-	-	+	-	-	+
AF Amharic	1.0	+	-	-	+	+	+	+	0	+	-	+	-	-	-	+	+	+	-	-	+
NG Yimas	0.5	+	-	-	+	+	+	+	+	+	+	+	+	-	-	-	+	-	+	-	N
NA Squamish	0.5	+	+	+	+	0	+	+	+	+	-	-	+	-	-	-	-	+	-	-	-
AF Lango	0.	+	+	+	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	+
SA Hixkaryana	-0.5	+	-	-	+	+	+	+	0	+	-	-	-	+	+	+	0	+	-	-	-
NG Mangarayi	-0.5	+	-	+	+	+	+	-	0	+	-	-	-	+	-	-	-	+	-	-	+
NE Akkadian	-0.5	+	-	+	+	+	+	-	0	+	-	-	-	+	-	-	-	+	-	-	+
NE Abkhaz	-0.5	+	-	-	+	0	-	+	+	+	-	+	-	+	+	+	+	-	-	-	-
NE Persian	-1.0	+	-	+	+	+	+	-	-	+	-	-	-	+	-	-	+	-	+	-	-
AF Wolof	-1.0	-	+	+	+	+	+	-	-	+	+	0	+	-	-	-	-	-	-	+	+
AF Gbeya	-1.5	0	+	+	+	+	+	+	-	-	-	0	-	-	+	-	-	+	-	-	-
NG Hua	-2.0	-	-	+	0	+	+	+	0	+	0	-	-	0	+	+	+	+	-	N	-
NA Miwok (Lake)	-2.0	-	-	+	+	+	+	+	+	-	-	-	-	+	-	+	+	+	-	-	-
EA Gilyak	-2.0	+	-	-	+	+	+	+	+	-	-	-	+	0	0	-	-	+	+	+	-
AF Coptic	-2.0	+	+	+	+	+	+	-	0	+	+	-	-	+	-	0	-	-	-	-	-
NA Cree	-2.5	-	+	0	+	+	+	+	0	+	+	-	-	-	-	-	-	-	-	-	-
EA Lahu	-2.5	-	-	+	+	+	+	+	-	-	-	-	-	0	+	+	+	+	-	+	+
SA Tzutujil	-3.0	+	+	+	+	0	-	+	-	+	-	-	-	0	-	0	+	-	-	-	N
SA Mixtec	-3.0	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
NG Dyirbal	-3.0	-	-	+	+	+	+	+	+	-	-	-	-	-	-	-	-	+	-	-	N
NA Eskimo	-3.0	+	-	-	+	0	0	0	0	+	-	-	+	+	+	-	-	+	-	-	-
SA Yagua	-3.5	+	+	0	+	0	0	-	+	-	-	-	+	-	-	-	-	-	-	-	N
NE Sumerian	-3.5	-	-	+	+	+	+	-	+	+	+	+	0	-	-	-	+	-	-	-	-
NA Slave	-3.5	+	-	-	+	+	+	-	-	+	-	+	+	-	-	-	-	-	-	-	-
NA Kiowa	-3.5	-	-	-	0	+	+	-	-	+	-	-	-	-	-	-	+	+	-	-	N
AF Yoruba	-3.5	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	+	-	-	-	+
SA Pipil	-4.0	+	+	+	+	+	+	+	-	+	-	-	-	-	-	0	-	-	-	-	-
OC Fijian	-4.0	-	+	+	+	N	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
NA Maricopa	-4.0	-	-	+	+	+	+	+	+	+	-	-	-	-	-	0	-	-	-	-	-
NA Koasati	-4.0	+	-	-	+	+	+	+	+	+	-	0	-	-	-	-	-	+	+	+	N
EU Basque	-4.0	-	-	-	+	+	+	-	0	+	-	-	-	-	-	+	+	+	+	-	-
EA Chrau	-4.0	-	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	N
AF Nubian	-4.0	+	-	0	+	+	+	+	+	-	-	-	0	+	+	0	-	-	-	-	-
AF Afar	-4.0	-	-	-	+	+	+	+	+	-	-	-	+	-	-	+	+	-	-	-	N

Table 8: Clusterings of + and - scores (continues next page)

		1	2a	2b	3	4a	4b	4c	5	6	7	8	9	10	11	12	13	14	15	16	17
OC Tagalog	-4.5		+	+	+				-	-			+	-	-	-	-	-	-	-	+
EA Chinese	-4.5	-	+	-	+	+	-	-	-	-					-	+	-	-	-	-	+
EA Burushaski	-4.5	+	-	-	+	+			-	+	0	-	-	-	-	-	0	+	+	-	-
AF Mandinka	-4.5	-	0	-		0	0	-							-	+	+	+	+	-	-
AF Maasai	-4.5	-	+	+	+	-			+	0	-	-	-	0	-	-	-	-	+	-	N
NG Tauya	-5.0		-	-	+	0			-	+	-		-	-	0		-		+	-	N
AF Nkore-Kiga	-5.0	+	+	+	+	+	-	+	+	+	-	N	-	-	-	-	-	-	-	-	-
AF Babungo	-5.0	-	+	+	+	+	+	-	-	-		N		-	-	-	-	-	-	-	+
OC Hawaiian	-6.0	-	+	+	+	+	-	+	+			-		-	-	-	-	+	-	-	-
NE Turkish	-7.0	+	-	-	+	0	-	-	0	-			-	-	-	-	+	0	-	+	-
AF Hottentot	-7.0	-	-	-	+	+	+	-	0	+	-			-	-	-	-	-	+	-	-
NG Amele	-7.5	-	-	0	+	+	-	+	-	+	-		-	-	-	-	+		-	-	-
EA Tamil	-7.5	-	-	-	+	0	-	-	0	-			-	-	+	-	0	-	+	-	-
EA Japanese	-7.5	-	-	-	+	+	-	-	-	-			-	-	-	+	0	-	+	-	-
NG Kobon	-9.0	-	-	-	+	-	-	-	-	-			-	-	+	-	+	-	+	-	-
NE Ingush	-9.0	-	-	-	+	-	-	-	+	-	-		-	-	-		0		-	-	-
EU Hungarian	-9.5	+		-	-	0	-	-	-	0	-	0	-	0	-	-	-	-	-	-	+
EU Albanian	-9.5	-	+	+	0	-	-	-	-	+	+	-	-	-	-	-	-	+	-	-	-
EU French	-10.0	-	+	+	-	-	-	-	-	-	+	+	-	-	-	-	+	-	-	-	-
EU English	-10.5	-	+	+	0	-	-	-	-	-		+	-	-	-	-	-	-	-	-	-
NA Shoshone	-11.5	-	-	-	+		+	-	0	-	-		-	-	-	-	-	-	-	-	-
SA Quechua	-12.0	-	-	-	+		-	-	0	-	-		-	-	-	-	-	-	-	-	-
NE Hittite	-12.0	+	-	-		0	0	-	-	-			-	-	-	-	-	-	-	-	-
NE Georgian	-13.0	-	-	-	-	-	-	-	-	+	-		-	+	-	-	-	-	-	-	-
EU Greek (Cl)	-13.5	-	-	+	-	-	-	-	-	-		-	-	-	-	-	-	+	-	-	-

Clusterings of + and - scores in the sorted data set
Table 8

and - polarities (ignoring quantitative value), with the languages rank-ordered as in Table 2a. In this format it is possible to eyeball the table and get an informal assessment of the degree to which + scores for any given feature tend to cluster in the upper half of the ranking.

Of course, some correlation exists and must exist, since a "good" CHS-type language is usually of VSO macrotype and fairly strongly Head-marking. A glance at the top 14 "best" languages in Table 8 (recall sec. 6.2) will reveal a strong dominance of + scores for most features. Yet even within these "best" languages, we see that none of the type-defining features must come up positive:

Clause-level WO:	SOV (Abkhaz, Akkadian, Amharic, Persian); SVO (Hausa, Lango, Wolof)
NP-level WO:	largely Dept-Head (Abkhaz, Amharic, Yimas)
Conjugated Adp:	not found (Wolof)
Polypersonal Verb:	only sometimes in Welsh; in Egyptian, Object pronouns are non-rigidly bound clitics

Indeed, for almost every feature in the feature set, this top group of languages shows a - value somewhere; exceptions are feature [3] (Special Relative Verb), which almost never scores negative anywhere, and [4a] (Copying in Gen RCl).

That the + values will cluster heavily in these top 14 languages is to be expected. What is the picture if we ignore these languages and look at the bottom four-fifths of the distribution? There are several cases where verb-first languages appear quite low in the ranking (Tagalog, Maasai, Hawaiian, modern spoken French); SVO languages may also

score low (Babungo, Nkore-Kiga, Albanian, English). With regard to Head-marking, again languages showing Head-marking features can be observed near the bottom of the ranking: Turkish, Hungarian, and Hittite have conjugated adpositions, while Albanian, French, and Georgian have a polypersonal verb. However, the overall trend is for the word order features [2a,b] to show a very weak "upward clustering" of + scores; as for Head-marking, feature [1] shows a somewhat stronger upward clustering, while [6] is almost flat. Thus some overall correlation does exist between degree of CHS-ness and word order, and likewise between degree of CHS-ness and Head-marking. In light of the intercorrelations of equipollent features already noted (6.6.1), this should not come as a surprise.

6.6.3 Detailed assessment of selected correlations

Several of the feature-pair correlations are particularly interesting for what they tell us about the naturalness of certain clusters of features in the CHS languages. In this section we will examine three such cases.

6.6.3.1 Copying as an oblique relative-clause strategy

In most of the CHS languages, both types of oblique relative clause --- genitival and prepositional --- are formed in exactly the same way, via a resumptive pronoun. Intuitively this seems a very natural strategy, as does the fact that genitival and prepositional RCLs are treated identically. The two subtypes have the feel of being minor variants of

a single overarching type: "oblique".

When we look at the correlation between the Gap-Copy features for genitival [4a] and adpositional [4b] relative clauses, this intuitive perception requires modification. To be sure, the features do correlate rather well (75% same-polarity); of the same-polarity languages in the core sample, 7 score -- and 11 (including 4 CHS) score ++. But only 24 languages take part in this correlation at all; usually one or the other RCl feature has a 0, N, or I score. And in fact there is a striking asymmetry in the overall distribution of features [4a] and [4b]. For feature [4a], many more languages score + than - (20 vs. 8); for [4b] the preference is reversed (12 vs. 25), as can be seen from Table 3a. That is, languages with externally-headed RCl³² strongly prefer the copying strategy for genitival RCl_s, but gapping for adpositional RCl_s. In fact, an implicational universal holds within the data set, supported by the 33 languages having non-blank scores (+, -, 0) for both features: if a language does adpositional relativization by copying, then it must do genitival relativization by copying. Another interesting asymmetry is that the Gap/Copy type (score 0) is fairly common among the languages in the sample as a strategy for genitival RCl_s, but is not attested for adpositional RCl_s.

There are, moreover, interesting areal restrictions on the correlation between the two features. Of the 11 languages scoring ++, 4 are CHS, 4 come from Africa (Babungu, Gbeya, Hottentot, Lango), 2 from the Near East (Sumerian and Persian, the latter plausibly under Arabic

³² Recall that internally-headed RCl_s earn zero scores.

influence), and only one from elsewhere (Slave, in North America). Thus the seemingly "natural" symmetrical pattern seen in the CHS languages, whereby genitival and prepositional RCIs are both formed by Copying, is in fact rare outside the CHS languages and Africa.³³

6.6.3.2 Verb-first word order and nonconcord with verb

As we have seen, the phenomenon of verbal nonconcord (feature [9]) shows a weak correlation with VO word order [2a] (sec. 6.6.1). But the stronger link with verb-first word order is often asserted; thus Greenberg's "Universal" #33: "When number agreement between the noun and verb is suspended and the rule is based on order, the case is always one in which the verb precedes and the verb is in the singular" (1966a:94). Nonconcord is certainly not restricted to VSO languages in the database, but with one exception the languages are not in conflict with Greenberg's careful formulation of the universal. Excluding the CHS languages, 10 others have some form of nonconcord; 4 of these (2 SOV, 2 VSO) belong to the "flexible marking" type (score +1/2) which can mark plurality on N or V or both. Of the remaining 6 languages, 5 show no ordering restrictions on nonconcord: Lango and Wolof (SVO), Afar (SOV), Squamish (VSO), Yimas (free). Nonconcord is obligatory in Lango and apparently the norm in Afar; conceivably this might be an East African areal feature. In Wolof nonconcord occurs in certain verb inflections; in Squamish and Yimas it is simply optional.

³³ By contrast, 6 languages score + for feature [4a] but - for [4b]; all are Old World languages (including two African languages: Yoruba, Nkore-Kiga), but otherwise no areal pattern emerges.

It is a remarkable fact that the one non-CHS language showing any ordering restriction on nonconcord, Yagua, patterns inversely from the CHS languages. Yagua is described as VSO. Concord is marked as a proclitic on the verb. When a full-NP subject precedes the verb, there is no clitic; when it follows, the clitic does appear:

- (i) SubjNP \emptyset -Verb
- (ii) Clit-Verb SubjNP.

It is surely significant that the verb agreement marker here is a clitic, and specifically a proclitic, rather than a desinence. This suggests a way to explain away the exception to the universal: reanalyze Yagua as SVO, with the preverbal subject taken to be either a clitic or a full-NP; in (ii) the postverbal "SubjNP" would be recast as appositional. However, the clitics can show considerable phonological fusion with the verb, which makes it odd to treat them on a par with full-NPs; moreover, VSO (pattern [ii]) is the favored word order (see write-up in Appendix 2). These facts argue against the tentative reanalysis. Yagua would appear, then, to be a genuine counterexample to Greenberg's universal.

6.6.3.3 Predicative particle and adpositional periphrastic

Looking only at Welsh and Egyptian, it seems extremely natural to posit a principled connection between features [11] (Predicative particle) and [12] (Prepositional periphrastic). In both languages, a locational preposition is used to recast both predicative nouns and progressive verbs as adverbial adjuncts to a copula:

He is in a companion. / He is in going.

To the student of either language, the connection feels compelling and natural.

In light of this, the actual distribution of these two features comes as a surprise, for they show almost no correlation. Feature [11] earns a positive score for 12 languages. However, an unproblematic, solidly installed construction involving a true adpositional particle (score +1) occurs in only 7 languages, of which three (Berber, Egyptian, Welsh) are CHS. Of the remaining four (Gbeya, Hixkaryana, Mandinka, and Tamil), all but the first make use of an adposition meaning "as, like, in the capacity of"; only Gbeya uses a locative/instrumental Prep. --- For feature [12], positive scores occur in 14 languages, but eight of these are only weak matches (+1/2), involving not true adpositions but hedged adpositions, adposition-like particles, local case markers, or locational copulas. Of the 6 languages that do score +1, all use adpositions having local meanings ("in, at, on"), which of course makes excellent sense semantically; the languages in question are Egyptian, Irish, Welsh, and three non-CHS languages (Afar, Chinese, Mandinka).³⁴

Two conclusions will immediately be apparent. First, in only one non-CHS language do both constructions cooccur in their full-fledged form: Mandinka, a West African language. (Abkhaz and Hixkaryana have positive scores for both features, but not two +1 scores). Second, the two constructions seldom use the same adposition. Typically feature [11] is formed with an Adp meaning "as", while feature [12] uses a local

³⁴ Mandinka actually has two such constructions: one uses a Postp "on", the other a semantically complex Postp covering instrumental and purposive nuances.

Adp meaning "in, at".³⁵ The Welsh-Egyptian agreement now becomes truly startling, not just because of the numerous resemblant features but because these multiple points of resemblance are now seen to be highly exotic. That both languages have both constructions is itself exceptional; that they use local prepositions for both, and sometimes even the identical preposition "in", is unprecedented.³⁶ And the parallel goes further still: of all the languages in the sample, only in Celtic and Egyptian does feature [12] include the possibility of using different adpositions to indicate different tenses.

This triple parallel is exactly the kind of quirky evidence which, quite properly, merits a special place in arguments against coincidence. Yet caution is in order. For the Welsh predicative particle was probably not originally a preposition at all. In Old Welsh, the predicative particle is not the preposition yn "in" but a form int plausibly deriving from the old Celtic article (see T. A. Watkins 1962). The latter had merged with the Prep by Middle Welsh, though with different mutation behavior; and almost every phase in the development can be elucidated in purely Celtic-internal terms.³⁷ We are thus faced with a confusing surfeit of riches: both an internal and an external explanation exist, totally different in nature, each with its own compelling logic. Yet

³⁵ Exceptions are Tamil, where the Postp "as a" has a marginal use as an exponent of progressive semantics; and Gbeya, which uses a loc/instr preposition as a predicative particle. Significantly, the latter is a West African language.

³⁶ Welsh uses "in" in both constructions; Egyptian uses "in" in [11], either "on" or "in" in [12].

³⁷ I hope to demonstrate this in a future paper. To my knowledge, although the origin of predicative yn has been treated in the literature, the detailed course of the particle's diachronic evolution has never been laid out or analyzed.

the Celtic-internal explanation cannot alter the fact of the Welsh-Egyptian threefold agreement, nor the radically exotic nature of that agreement. Multiple causation might resolve the difficulty, with the Celtic account providing a plausible "how" and the substratal account a plausible "why". I return to this puzzle in sec. 7.4 below.

Chapter 7: Conclusions

7.1 Results

After such minute inspection of the terrain, we may finally be in a position to see both the forest and the trees. The results can be stated concisely. Roughly half the individual CHS features show the strongly skewed-exclusive distribution characteristic of typological exotics. This kind of skewed distribution is the best one possible in demonstrating that two distinct occurrences of the same phenomenon are not simply a matter of coincidence. In partaking of such rare features, the Celtic and Hamito-Semitic languages are "bucking a global trend" that strongly favors the opposite behavior, seven or eight times over. --- Some of the features show a degree of correlation among themselves which lessens their status as independent variables --- specifically, a cluster of equipollent Head-marking factors involving conjugated prepositions, polypersonal verb, relativization by copying, and verbal nouns. Word order does not appear to be part of this constellation, or only very weakly. However, correlations involving the exotic features are minimal, tend not to hold globally, and typically do not involve the full-fledged CHS phenomena (scores of +1/2, not +1). Significantly, Africa and the CHS languages contribute strongly to correlations involving exotics.

On the basis of the sample used in this study, nothing remotely close to the full-blown Celtic/Hamito-Semitic linguistic type recurs anywhere else in the world. The relatively few languages which are "best matches" --- actually rather poor matches --- are scattered all over the globe, from the West Coast of North America to the Caucasus and New Guinea. However, the continental average score for Africa is higher than for any other continent, and drops only slightly when the CHS languages Egyptian and Berber are omitted; West Africa scores especially well, and appears especially hospitable to several of the CHS features (adpositional periphrastic, word-initial change, kin terms, *inter alia*). Conversely, Europe has one of the lowest average scores, and when Welsh and Irish are excluded its score drops far below that of any other continent. Celtic is thus radically out of place in a European landscape, whereas the Hamito-Semitic languages simply intensify a structural trend seen over much of Africa. A weak form of the CHS type, then, would appear to have a natural home in Africa, in particular Northwest Africa. Within Afroasiatic, the highest-scoring languages are on the Mediterranean; scores fall away in every direction, but the Chadic language Hausa (in West Africa) scores much higher than Cushitic Afar (in East Africa). The diachronic evidence, too, argues that the (weak) CHS type is something quite old in Africa: the African and Arabian case studies all show stronger CHS-ness further back in time. All this, in conjunction with the blood-type agreement between the British Isles and Northwest Africa, argues for some sort of prehistoric scenario specifically linking these two regions. Conversely, Wagner's suggestion of a "Eurafrican area" founders on a lack of factual support elsewhere in Europe. Basque, the linchpin of his argument, scores poorly; Abkhaz,

with its startling agreement in blood type, scores high, but is poorly positioned geographically to be implicated in any link with Britain.

These are the concrete results of the survey. In the process of arriving at them, a host of issues have arisen, many of them prefigured in the general discussion in Chapter 1. The final sections of this study will follow up the implications of some of these general issues.

7.2 Two perspectives on substratal methodology

In this and the following section, I examine the CHS problem from two outside methodological perspectives. The first is that of archaeology, a field where diffusionist explanation of any sort has fallen heavily out of favor in the last two decades. The second is strictly linguistic, and examines the present study from the perspective of recent work by Thomason and Kaufman.

7.2.1 Archaeological antidiffusionism: The megaliths

The mystery of the megalithic tombs of western Europe, and the scholarly problem they represent, have already been mentioned briefly (sec. 1.5.3). Until quite recently scholars were in firm agreement that the megaliths represented a clear case of cultural diffusion westward from the Aegean (thus Daniel 1963). The argument rested partly on a methodological preference for a certain mode of historical explanation --- "ex Oriente lux", with culture radiating from the civilized Orient to the barbarians (Renfrew 1979:30ff.) --- but also on numerous points of structural resemblance between artifacts from East and West. With

the advent of radio-carbon dating, and in particular with the major adjustments to C14-dates resulting from tree-ring calibration, this notion was exploded: the megaliths were shown to be far earlier than the monuments of the eastern Mediterranean, and even than the pyramids (1979:123).¹ The scholarly consensus now tends to consider the megaliths cases of independent parallel development.² Yet for years diffusionist archaeologists had devoted earnest attention to adducing and analyzing points of typological resemblance between structures from the Mediterranean and the Atlantic. The worth of these similarities as evidence had been almost universally accepted; now the whole structure of the argument was exposed as illusory. And the debunking, which involved not only this but other concrete cases as well, could hardly help but carry a more general methodological corollary: an automatic skepticism regarding inferences of contact based purely on typological comparison.³ Scholars, and notably archaeologists and prehistorians, would be very cautious about making the same mistake twice, or about reapplying a method so dramatically exposed as fallible. And nowhere would this be truer than for an argument purporting to relink what had just been delinked: Western Europe and the Mediterranean.

The relevance of these considerations to the present study should be obvious. In both cases, typological similarities --- linguistic and

¹ The reasoning is laid out clearly and elegantly in Renfrew 1979; see also Renfrew (ed.) 1983.

² For a dissenting voice see MacKie 1977, who again advocates links to the Mediterranean.

³ Thus Renfrew: "The repeated rejection of theories based upon diffusionist principles has effectively called into question the validity of the traditional diffusionist approach in any field" (1979:2).

archaeological --- are appealed to as evidence in support of prehistoric contact. The archaeological argument has been exposed as invalid despite decades of acceptance in good faith by the scholarly community. Is it not hazardous, then, to propose exactly the same kind of argument in the linguistic domain?

In fact, though, the two cases are only weakly parallel, for the kinds of features involved differ fundamentally in several ways. In archaeology, the degree to which megalithic constructions in different lands are perceived as having striking similarities varies subjectively from scholar to scholar: Childe speaks of "strict agreements in arbitrary details of funerary architecture over large tracts of Mediterranean and Atlantic Europe" (quoted in Renfrew 1979:123), whereas Renfrew himself minimizes such similarities, presenting them as "not really very striking" (88) or simply ignoring them. By contrast, the CHS linguistic similarities involve a large number of exact or near-exact structural matches defined with a high degree of specificity. Secondly, megalithic structures are not unusual in the world, occurring in such diverse localities as Japan, India, Polynesia, and Mesoamerica (Daniel 1963:22-23). The CHS features do recur individually outside the CHS languages, but nearly half are exotics and as such globally rare. Finally, to a considerable degree cross-cultural commonalities in the forms of megalithic tombs can be seen as dictated by societal and functional considerations. A pre-metal society intent on building large tombs that would not be sealed off but allow multiple reentry, and inhabiting a region where the local rock resisted splitting, would have little option other than the use of megaliths (1979:129, 131-32). In linguistics, by contrast, this kind of elementary functional explanation

is excluded. To interpret features of a language's structure (as opposed to its vocabulary) as a response to the local environment or culture is a notion that appears to have no basis in reality. And in any event, one could hardly imagine two more different environments than Britain and western and northern Africa.

With the megaliths, the case for diffusionism was weak to begin with, and readily collapsed in the face of the radio-carbon revolution. Had the features of similarity been stronger; had the megaliths been rare or nonexistent elsewhere in the world; had there been no obvious functional motivation for the megalithic constructional technique --- then even with carbon dating it would have been much harder simply to shrug off all notion of contact. It is precisely these conditions, or their analogues, which hold for the CHS linguistic similarities.

7.2.2 Thomason and Kaufman on substratal explanation

One of the most important recent books on the theory and practice of historical inference in linguistics --- important both methodologically and for its wealth of carefully analyzed case studies --- is Thomason and Kaufman's Language contact, creolization, and genetic linguistics. This book presents a vigorous and sober defense of external explanation in historical linguistics, flatly rejecting the automatic bias in favor of internal explanation so prominent in orthodox diachronic methodology (lff., 57, 59, and passim). Hence it will be of interest to consider how the CHS problem, taken as a case study, at once harmonizes and clashes with the theoretical framework laid out in that book. In the process we will recapitulate many of the central

methodological points laid out over the course of this study.

Thomason and Kaufman mention the CHS problem only briefly, but what they do have to say is explicit and negative. On methodological grounds, they reject as speculative any attempt at substratal explanation for the structural changes undergone by Insular Celtic. A strict methodological prerequisite for any substratal explanation, they assert, is that the identity of the alleged substratum language or language group be established as a matter of fact, not of mere conjecture (63, 111). And "since we have no information about what language(s) the pre-IE inhabitants [of the British Isles] spoke, we cannot establish [a substratal] cause for these changes" (111-12). This methodological point is followed up by an assertion that "all the hypotheses that have been advanced about such a substratum ... rest on such tenuous historical and linguistic evidence that the chances for a convincing proposal in this area seem remote" (112).

By now it will be clear that the linguistic evidence is anything but tenuous. The authors are, however, quite correct in stating that there is next to nothing in the way of historical evidence for a posited Celtic/Hamito-Semitic connection. The substratal language, being unattested, is not historically identifiable. And they draw the traditional methodological conclusion: in this domain, at least, such direct historical evidence is a sine qua non. That is, in the sphere of substratal influence, purely linguistic evidence by the very nature of things can never provide a sufficient warrant for nonspeculative projections into prehistory. Indeed, at first glance the point would appear obvious.

This thinking reflects an odd blind spot. For there is one domain where we routinely do make projections into linguistic prehistory based purely on linguistic data: genetic linguistics. Proto-Indo-European is not an attested language, nor is there to my knowledge anything beyond purely linguistic data to link (say) Algonquian and Wiyot/Yurok --- yet here we speak of proven relationships, and proceed with confidence to articulate details of reconstructed prehistory. Here, far from grounding our linguistic deductions in attested facts of prehistory, we instead boldly put forward a view of prehistory based purely on linguistic data.⁴ We do this because we feel there is "compelling crosslinguistic evidence" --- genetic evidence, in this case, applied to a genetic problem (the reconstruction of protolanguages and of a family's genetic history). And we can appeal to the tried-and-true comparative method as a warrant for the reliability of our abstract projection back into prehistory.

There is in principle no reason that the same kind of thing, buttressed by a different kind of historical method, could not be possible in a nongenetic diachronic arena. This study has presented such a method, which provides a way to assess the probative value of structural resemblances among unrelated languages in order to venture probable projections into prehistory even in the absence of concrete, real data from that unknown period (here, the linguistic situation in pre-Celtic

⁴ Of course, scholars do attempt to articulate concrete details of IE prehistory and to uncover archaeological correlates for prehistoric Indo-European speech communities. But these are subsequent to, not a prerequisite for, the purely linguistic process of inference that established the notion of Indo-European family and protolanguage in the first place.

Britain). Of course, this method cannot attain to the reliability of conclusions based on the comparative method. But it does provide a non-speculative way of drawing probable inferences about prehistory. Indeed, it is easy to imagine a scenario where even the most skeptical would be convinced --- a scenario where Celtic and Mediterranean Hamito-Semitic had (say) 50, or 100, independent exotic features in common. At some point, to continue to insist on concrete historical knowledge becomes absurd: the cumulative weight of the evidence ultimately builds toward a "proof" analogous to that provided by the comparative method. It simply makes no sense to insist as a matter of logic that, in the absence of concrete (pre)historical data, genetic projections into prehistory may aspire to legitimacy while nongenetic diachronic projections must be illegitimate. Nothing privileges genetic linguistics in this respect except its possession of a reliable method for drawing prehistoric inferences. When such a method is developed for other subdomains within diachronic linguistics, the need for concrete historical corroboration becomes correspondingly less important.

The "historical identifiability" criterion just discussed is the sole point where the arguments presented in this study are at loggerheads with those adduced by Thomason and Kaufman. In other respects, the CHS problem would seem a textbook case for positing substratal action, even in Thomason and Kaufman's terms. To make a good case for substratal action, they argue,

we must be able to identify a substratum language or language group (some of) whose speakers shifted to the target language at the relevant time period; we must have information about

its structure; and we must have information about the structure of the target language before the shift. (111)

As remarked, in the CHS case we cannot actually identify the substratal language; but we do know that some substratal language did predate the Celts in the British Isles, and the typological considerations laid out in Chapter 6 argue strongly for connections with North Africa. We do have information about the structure of this substratal language, assuming it to have had North African connections: its structure cannot have been radically different from that of Berber, Egyptian, or Semitic. And we have information about the structure of pre-Insular Celtic prior to any type shift, both from Continental Celtic inscriptions and from our knowledge of Indo-European.

Other aspects of the CHS problem harmonize well with Thomason and Kaufman's discussion of substrata. First, substratal change (as opposed to structural borrowing) can easily involve little or no transfer of vocabulary (39); and indeed there are no evident lexical correspondences between Celtic and Hamito-Semitic (recall sec. 1.5.5). Second, although Thomason and Kaufman point out that corresponding features in the substratal and the affected language need not be exact structural matches (61ff.), in fact most of the CHS features do show up in near-identical form in Celtic and Hamito-Semitic --- a fact which can only strengthen the case for substratal action. Third, in order to infer substratal (as opposed to borrowing) interference, Thomason and Kaufman make the methodological demand that features of similarity not be confined to a single linguistic subsystem (60). Indeed, the CHS features do encompass more than one subsystem --- primarily morphology and syntax, but also

impinging on phonology (word-initial change) and even on the structure of the lexicon (kin terms). One reason put forward for insisting on multiple linguistic subsystems is as a corrective to the possibility of putatively independent features actually turning out to be correlated (347-48, note 9). We have addressed this issue of feature correlation directly. Fourth, the authors note (following Heath) that the sharing of a typologically rare category has special probative value in an argument for substratal action, "as a methodological strategy for convincing skeptics" (60). We have remarked that nearly half the CHS features are typologically rare exotics. Finally, in cases where a language and its alleged substratum share numerous independent resemblances, Thomason and Kaufman note that invoking a substratal explanation (as opposed to piecemeal family-internal explanations) has the cardinal virtue of providing a unified account instead of a scattering of atomistic accounts (61). The CHS languages share a very large number of resemblances (twenty), thus rendering a substratal account all the more highly valued.

7.3 Criticisms and responses

Over the course of almost a hundred years of investigation, many methodological and theoretical objections to a substratal account of the CHS problem have come up; in this section they will be reviewed and treated as a body. The typological method developed in the preceding chapters was designed with an eye to neutralizing the most fundamental of these objections. Thus Pokorny and Wagner were severely criticized for their dogmatic faith in the universal explanatory efficacy of

substrata or areality. They neglected the very real possibility of coincidence or typological affinity among features. And their methodology --- impressionistic, speculative, uncritical --- seldom transcended the slogan "Es ist wohl kein Zufall". These criticisms are completely valid; but they do not apply to the typological method.

It is often pointed out that certain individual CHS features are derivable from Indo-European, thus "obviating" the need for an external, substratal, explanation. Several answers can be made to this objection. First, the notion that one is entitled to propose either internal or external explanation but not both is methodological purism carried to a pointless extreme. Multiple causation is hardly a new idea in linguistics; and when the competing explanations are internal (genetic) vs. external (contact), an obvious and attractive synthesis suggests itself. A genetic explanation traces a plausible path whereby the grammatical material of the parent language could metamorphose into the phenomenon seen in the descendant, but it seldom offers a reason why this path of development, of all possible paths, should have been the one followed. An external explanation explicitly addresses this point, but typically says little about the details of the evolution. The two forms of explanation thus answer different questions. Internal explanation provides a plausible "how", and external explanation a plausible "why".

The above objection also carries the methodological presupposition that internal explanation should automatically take priority over external (on this see Thomason and Kaufman 1988:57ff.). On this view it is only with radically new phenomena, inexplicable in internal terms and diverging sharply from the earlier course of a language's development,

that external explanation is entitled to consideration. But this has been recognized since Brugmann as fallacious. More typically, contact will serve to foster some recessive option already present in an early form of the target language, raising it to new prominence; this option will be privileged precisely because it is a good structural match for an equivalent phenomenon in the substratal language (see secs. 1.5.4 and 2.3.1.2). Thus the fact that a given phenomenon has an internal, genetic derivation can be fully compatible with a substratal account.

Another response is that suggested by Thomason and Kaufman (1988:61) and noted in sec. 7.2.2. IE-internal explanations will of necessity be atomistic; at best several factors (notably the development of a polypersonal verb, the infixing/suffixing alternation, and certain details of the RCl) may cohere in a single explanation. By contrast, a successful substratal account treats the problem as a unity.

There is, finally, a rather different approach that can be taken to the "internal derivation" objection. It must not be forgotten that the CHS problem actually has two prongs: a radical disagreement with Indo-European, and a radical agreement with Mediterranean Hamito-Semitic. It is in fact two subproblems collapsed into one, and each represents a bona fide issue in its own right. The above objection embodies a presupposition that the primary focus is and should be on the disagreement with Indo-European --- the internal, genetic question. Looking primarily at this question will naturally foster genetic analysis, genetic answers, a genetic way of thinking about the problem: one will try to show how the Celtic phenomena do (after all) emerge from configurations within IE. Similarities to HS, on this view, have no

intrinsic interest; they enter the problem only secondarily and incidentally, when no internal explanation is forthcoming. But in fact the second problem, the massive agreement with HS, is fully as legitimate as the first, though historical linguistic tradition is not used to thinking of it in these terms. One can choose to focus on either question; and if one focuses on the second, then the analysis, answers, and mode of thinking become typological, and it is internal genetic explanation that becomes secondary. And because each question is legitimate in its own right, its legitimacy is independent of anything we may know about the other. Obviously the problem of Celtic disagreement with IE would be just as valid if there were no HS languages; similarly, the problem of agreement with HS would be just as valid if there were no IE superfamily at all and Celtic were an isolate, or (conversely) if every aspect of the IE-internal subproblem had been worked out in full. IE-internal explanation, no matter how insightful, simply looks right past half the problem.

A second recurrent methodological objection is that no substratal explanation of the CHS problem can ever be convincing simply because we lack all direct knowledge of the pre-Celtic language(s) of the British Isles. As argued at length in sec. 7.2.2, this common-sense view is deeply fallacious. A lack of direct knowledge of the relevant prehistoric (proto)language is never an obstacle in genetic linguistics, where we routinely reconstruct protolanguages based purely on linguistic data; nor should it be an obstacle in positing unattested substrates, if a solid method is in hand to provide a warrant for such prehistoric projections. In any event, in this case some nonlinguistic evidence does exist: the blood type data cited in sec. 1.5.3 does provide a degree of

extralinguistic corroboration for a connection with Northwest Africa, as does the Barbary ape skull.

In a different vein, it can be objected that certain CHS features do recur sporadically elsewhere in older Indo-European. Indeed they do, just as they recur sporadically in languages all over the world. The point is not any individual feature, but the cumulative weight of the ensemble. One might also point out that few of the CHS features occur in all the CHS languages; features [4b] and [4c] (adpositional RCl strategy) are in fact close to mutually exclusive. But to insist on this point would be unreasonably stringent. In neither genetic nor areal linguistics do we demand that a characteristic feature or etymon be found in every language of the family or area. A few such items will happen to occur throughout the family or area; similarly, a few CHS features (e.g. Conjugated prepositions, or NP-level word order) hold everywhere in the group. But this is the exception, not the rule.

Critics sometimes point to the absence of lexical resemblances between Celtic and Hamito-Semitic. As Thomason & Kaufman explain (39), this is quite typical of substratal situations. But even if vocabulary resemblances were to be expected on theoretical grounds, in practical terms we would be unlikely to find any (recall sec. 1.5.5). The particular substratal language is of course unknown; millennia have elapsed since the time of putative contact; and both Celtic and HS are notorious for having undergone major phonological changes, effectively disguising even those few words that might otherwise have survived in recognizable form.

There is, however, a potential objection of a much deeper sort. Surely there are other instances where two languages or language groups share a large battery of structural similarities. Indeed, one might think that, with a little effort, such a set of multiple resemblances --- involving different features, of course --- would stand a good chance of being found in many language pairs selected at random. Perhaps the present study is seizing on one such set arbitrarily, and misrepresenting it as something special.⁵

Four responses can be given. The first is to emphasize once again the importance of the skewed-exclusive nature of nearly half the features. These features are exotics, phenomena involving unusual possibilities of language structure and bucking global trends to the contrary. One cannot prove such things, but I would suggest that multiple resemblances of this sort are extremely unlikely to be found between languages selected at random. The burden of proof is properly on those who propose the contrary.

Second, the above argument would seem unlikely to have been raised had the number of exotic resemblances between Celtic and Hamito-Semitic been higher (say, fifty). There are no quantitative canons for how many resemblances of this sort are needed to forestall objections, but at some point the resemblances become so numerous that it is hard to

⁵ The issue transcends linguistics, of course. Renfrew, writing in an archaeological context, points out the "circularity inherent in many diffusionist arguments, where features in different cultures, often widely distributed in space and time, are selected for comparison because they do show similarities. It is then held to be significant that the features thus selected are indeed so similar. In reality, however, much of the force of this argument comes from the exclusion of the other evidence which could suggest very different conclusions" (1979:126).

imagine a similarly large set occurring routinely with other language pairs. Eight-odd exotics, plus a dozen more "normal" resemblances, appears to me quite a large quantity; recall, too, that the CHS resemblances outnumber the set of features defining the Balkan Sprachbund by two to one.

Perhaps most important, however, is the third argument: the fact that the set of CHS similarities have a double warrant. Not only do they constitute a set of resemblances between Insular Celtic and Mediterranean Hamito-Semitic, but at the same time a set of startling antiresemblances between Celtic and its own parent family, Indo-European. Above I emphasized that these represent two independent issues, which can and do come up singly for other sets of languages all over the world. One of the most remarkable aspects of the CHS problem is that here both considerations are copresent and indeed almost congruent. This will clearly not be the case for sets of resemblances (or antiresemblances) between random languages.

Finally, if the kind of scenario proposed did occur --- if two other languages or language groups were to share a remarkable number of features showing a strong skewed-exclusive distribution --- I believe that the same conclusion arrived at in this study should make equally good sense there. Nothing will have been proven, but coincidence will have been shown to be very unlikely. If there is any remote possibility of a (pre)historical account involving contact between the groups, it would merit very serious consideration.

There remains one major criticism which has barely been touched on so far. The substratal action posited in this study has in effect been operating in a temporal vacuum, "sometime in prehistory". In part this is inevitable, given the nature of the problem; and often it is harmless. However, certain CHS features show anomalies of absolute and relative chronology that cannot just be ignored: the features show up at the wrong stage of a language's history, too early or too late to be straightforwardly incorporated in a substratal scenario. This large and complex issue is the theme of the next section.

7.4 Chronological anomalies in a substratal account

In examining the CHS problem from a diachronic perspective in sec. 6.5, we mentioned the fact that, although Irish in overall profile does not become more CHS-like over time, nonetheless individual CHS features absent in Old Irish do appear in Modern Irish. Several other such chronological difficulties exist as well. A few CHS features found in Middle Welsh and Old Irish are absent or marginal in the very oldest stages of the languages (Old Welsh, Archaic Old Irish). Some of the CHS features can already be seen in Continental Gaulish. And in one case, a Berber phenomenon adduced as a possible source for its Celtic analogue may itself be historically secondary. All these chronological anomalies are potentially damaging to any attempt at substratal explanation. We need to consider, therefore, whether there is some way to reconcile these puzzling facts with a substratal account. The proposals in this section are speculative but reasonable; I do not advance them as definitive solutions, but rather to show that there is no logical barrier to

resolving these temporal incongruities within a substratal framework.

It must be emphasized, first of all, that most of these objections are peripheral to the central focus of this study, which is Mediterranean Hamito-Semitic and the plausibility of a prehistoric link between Britain and Africa. The chronological anomalies have little to do with Hamito-Semitic or with any specific substratum; rather, they concern the workings and applicability of pre-Celtic substratal explanation per se. The crux arises from the fact that substratal action on Insular Celtic has only a limited temporal and geographical window within which it should have operated. First, the substratum is assumed to have run its course in the old phases of Insular Celtic (recall sec. 3.1); hence CHS phenomena that first appear in the modern languages (Modern Irish) cannot straightforwardly be explained in substratal terms. Second, CHS phenomena existing in the "classical" stages of Insular Celtic (Old Irish, Middle Welsh) but not yet well-installed in the preclassical languages (Archaic Old Irish, Old Welsh) likewise strain the chronology.⁶ In such cases a substratal explanation is forced to posit the continuing existence and influence of the substratal language at a very late date, postdating the preclassical languages. Finally, the substratum is assumed to have existed specifically on the British Isles; hence prior manifestations of the CHS type in Gaulish come outside its purview, demonstrating that these phenomena could and did arise in early Celtic even without substratal influence.

⁶ Recall (sec. 3.1) that these preclassical stages are not used as the basis of description for Welsh and Irish in this study; the corpora are not big enough, and the grammars not always well enough understood.

These three types of Celtic anomalies represent two quite distinct lines of attack on substratal explanation. On the one hand, the Old Welsh and Archaic Old Irish cases have direct and tangible consequences for a substratal account of the classical languages, and in particular for its chronology. If substratal explanation is to account for a phenomenon which is absent in an attested preclassical language but present in the later classical form, then the substratal language must still have been spoken as late as the early centuries of Celtic recorded history. Not only must it have been spoken, but it must still have been sociolinguistically salient enough to exert structural influence over the development of Celtic. This in effect forces the continuing presence of a large non-Celtic-speaking speech community at an improbably late date. By contrast, the Gaulish and Modern Irish anomalies have no direct impact on a substratal account of CHS-ness in Old Irish and Middle Welsh. Even if no explanation at all can be found for these anomalies, the worst they can do is to show that isolated CHS features can arise in Celtic without substratal influence. This may reduce somewhat the abstract plausibility of invoking a substratal scenario for Old Irish and Middle Welsh, but it will not impinge in any concrete way on the workings of such a scenario.

Clearly, then, the Old Welsh and Archaic Old Irish anomalies constitute the more urgent problem. Failure to explain these chronological anomalies leads to results verging on a reductio ad absurdum; failure to explain the Modern Irish and Gaulish anomalies merely leads to a lowering of plausibility.

In a different way, however, the two types of strictly Insular

anomalies --- that posed by Modern Irish, and that posed by Old Welsh and Archaic Old Irish --- are actually quite similar. They represent two manifestations of a single problem: operation of the substratum at an unreasonably late date. The same general types of explanation, therefore, are in principle relevant in both cases. First, it might be proposed that the substratum did in fact survive until the "unreasonably late date". This explanation, barely tenable for the Old Welsh and Archaic Old Irish problem, seems completely impossible for Modern Irish (recall the argument presented in sec. 3.1). Second, a multi-dialectal account can be suggested. Another dialect or dialects may well have coexisted alongside that represented by the written language; the latter, conservative and literarily self-conscious, might adhere to an older standard long after other dialects had changed under substratal pressure. In this way, even after the substratal language died out, its structural influence could live on in the form of nonwritten dialects, eventually to penetrate the written standard centuries later. This is in essence Pokorny's argument (sec. 2.3.1.2). And third, it may be that natural typological affinity is at work (sec. 6.6.1).

There is also another option, at least in the case of Modern Irish and Gaulish (representing the less urgent type of anomaly): there may be no compelling need to explain the anomalous change at all. Languages routinely show a change of some isolated feature for no discernible reason, and Celtic should be no exception.⁷ Here the non-Celtic CHS

⁷ Recall that the reason for invoking substratal explanation in the case of Old Irish and Middle Welsh is not the presence of one or a few CHS features, but a sweeping, multi-feature change in overall type (away from Indo-European and toward Mediterranean Hamito-Semitic). Such a characterization certainly does not apply to the changes observed between Old and Modern Irish, or between IE and Gaulish.

languages provide a useful concrete parallel. We have seen that these languages can alter their feature profile over time, sometimes involving an increase in the CHS-ness of individual features (recall sec. 6.5 and Table 6) --- just as Modern Irish does vis-a-vis Old Irish (and Gaulish vis-a-vis its parent Indo-European). Particularly instructive is the case of Amharic, whose development (as remarked in sec. 6.5) is similar to that of Irish in that the feature profile shifts but the overall score does not. Like Irish, Amharic acquires several new + scores vis-a-vis its ancestor language; and not all can be attributed to substratal influence. Yet this could hardly be said to vitiate the reality of Cushitic substratal effects on Ethiopic Semitic. In the same way, there is no reason to insist on substratal explanation for scattered developments of the same sort in Modern Irish (or Gaulish). And there certainly is no reason to demand that substratal explanation, in order to legitimately apply in the case of Old Irish and Middle Welsh, must also apply to every attested instance of increase in a CHS feature in every Celtic language. --- In fact, plausible explanations will be proposed for every case of anomaly; but rejecting these speculative suggestions does not ipso facto annul the validity of substratal explanation.

We turn now to a survey of the individual chronological anomalies, beginning with Modern Irish. Here a multi-dialectal explanation seems the most promising. As remarked in sec. 6.5, Irish acquires the following three CHS features in the post-Old Irish period:

- [4b] Adp RC1 by Gapping (rather than by Move-Adp)
- [9] Verbal nonconcord
- [11] Predicative particle ("He is in his doctor")

Feature [4b], however, is already present in traces in Old Irish. It has been tentatively explained as representing the survival of an old geographically based dialect split already existing in Old Irish times (McCone 1985:97). What we know as Old Irish, McCone argues, represents a northern dialect, in which relativization followed schema [4c] (as even today in Scots Gaelic); it was the old southern dialect, rather, which had prepositional relativization according to schema [4b].

As best I know, the concrete evidence for construction [4b], and hence for positing its existence in an ancient dialect, consists literally of a handful of forms (two examples in Thurneysen 1946:322, another alluded to in Ahlqvist 1988:35, note 7). Yet the case seems plausible despite the meager evidence --- spoken Old Irish surely was not as structurally uniform as would appear from the corpus of the glosses --- and scholars such as McCone and Ahlqvist endorse it. What the paucity of evidence demonstrates is how little such dialects might succeed in penetrating the structure of the standard language. It is thus not unreasonable to guess that some early Irish dialect(s) L might have had nonstandard feature(s) which completely failed to penetrate canonical Old Irish. And such features might have included versions of [9] and [11], which in dialect(s) L might then be assumed to have arisen under substratal influence. Such dialect(s) would have subsequently contributed these constructions to later forms of Irish.

Obviously this is a guess. I advance it here simply because it offers a multi-dialectal mode of explanation which is not at all implausible and which is completely in harmony with a substratal account. For example, Dillon describes the evolution of feature [11] ("He is in his

doctor") in Irish, where it makes its first appearance in Early Middle Irish (1927-28: 324ff.) as a natural extension of earlier constructions; conceivably that evolution might have already been underway centuries earlier in a spoken dialect. --- On the other hand, with regard to feature [9] (Nonconcord), I mentioned in sec. 6.5 the weak possibility of some natural typological correlation of feature [9] with certain of the CHS features. In that case the appearance of feature [9] in Modern Irish would not need external motivation. Only a separate typological research project can establish how valid such a correlation really is, and hence how much explanatory force it merits.

It is curious that the same pair of features, [9] and [11], also muddies the early history of Welsh. Here the problem is not so much the late development of the phenomena, as the peculiar light cast on them by the limited corpus of Old Welsh. Regarding feature [9], first of all, Old Welsh presents a handful of examples showing full concord in V-S order, and a handful showing nonconcord (D. S. Evans 1971:43, 49). This differs from the norm in Middle Welsh, where nonconcord dominates strongly and where (so it is argued [Evans 1971]) the occurrence of concord represents an attempt to imitate Latin literary models. Yet it seems clear that nonconcord, albeit only as an option, already did exist in Old Welsh and likewise in Old Breton (Fleuriot 1965:415), plausibly pointing to an old Brythonic (though not an Old Irish) innovation vis-a-vis Indo-European. This is certainly not incompatible with a substratal account; Goidelic and Brythonic need not have reacted identically to substratal pressure. Indeed, it would seem very natural to explain the persistence of some degree of concord as primarily a conservative

literary manifestation (Evans 1971:49).

Feature [11] (Predicative particle) presents a more serious difficulty. In Old Welsh, unlike Middle Welsh, the predicative particle is spelled int, differently from the Prep in "in", and hence the two cannot plausibly be taken as identical. Apparently int, originally a petrified article, does not become in until the 9th century AD (T. A. Watkins 1962:299). Now it will be recalled (sec. 6.6.3.3) that there is extremely strong circumstantial evidence in favor of linking the Welsh and the Egyptian phenomena here. Yet if an Egyptian-like language had a hand in the development of the Welsh construction, the Welsh predicative particle arguably should have been a preposition (as in Egyptian) even in Old Welsh.

One response is to suggest that the pre-Brythonic substratal language (whatever its ultimate origin) was still spoken and still exerting influence on Welsh in the 9th century AD. Is this possible at such a late date? Conceivably. It was only in the 9th century that the Picts lost their independence (Wainwright 1955:4); plausibly, therefore, the Pictish language was still spoken at that period (see discussion in sec. 3.1). If we assume, again plausibly, that Pictish was non-Indo-European, and was identical to or at least related to the pre-Welsh substratal language, then we have weak evidence suggesting that a substratal language may still have been in place and in use in the 9th century. But it seems improbable that such a speech community might have still been numerous enough to exert any kind of substratal influence on Celtic at such a late date.

A somewhat different approach to the problem could be taken, however. We have been focusing on the prepositional nature of the predicative particle; one could instead focus on the fact that a predicative particle exists at all. Indeed it does exist in Old Welsh; and, although the same etymon int appears throughout early Celtic (including Welsh) as an adverbial particle, the predicative use seems unique to Welsh (see T. A. Watkins 1962:300). Functionally, therefore, the phenomenon cannot be considered an inherited Celtic feature. This aspect of the problem is indeed consonant with a substratal account. And the forms int and in, even if not identical, are similar. One could imagine bilingual speakers, accustomed in their non-Celtic language to using a preposition "in" both as predicative particle [11] and in a prepositional periphrastic construction [12], seeking to extend this speech habit as best they could to pre-Welsh. They would of course have been aware of the phonetic similarity between the Welsh preposition in and the adverbial particle int, a similarity which, under substratal influence, could have led to functional association of the two particles even prior to later merger. --- Note that this account again requires a relatively late substratal presence, since the predicative use did not develop elsewhere in Brythonic. But the lateness is not as extreme as the 9th century. On this account, the substratal language would have had to still be spoken and widespread at the time when Welsh emerged as a separate language, which occurred by the end of the 6th century AD (Evans 1976:xvi). As far as I know, nothing contradicts such a view; Pictish surely was still vigorously spoken at this period.⁸

⁸ The Venerable Bede, who died in the 8th century, mentions Pictish in his list of languages of Britain (Wainwright 1955:3).

An analogous problem exists in Old Irish with regard to the article in genitive constructions (feature [8]). The Semitic-like placement rule, well established in Old Irish, is not yet a feature of Archaic Old Irish. This is not because some different rule existed, but simply because the category "Article" itself is absent. If the Irish rule came into being under substratal influence, then the substratal language must still have had to be widely spoken in Ireland subsequent to Archaic Old Irish. Presumably this would point to a period between the 6th and 8th centuries AD (Thurneysen 1946:4, 8-9; Greene 1977:14). One may wonder whether this is unreasonably late to posit a continuing substratal presence in Ireland. As far as I know it is not impossible. But two alternative explanations may be suggested.

"Archaic Old Irish" is fundamentally not so much a chronological label as a stylistic and generic one, designating primarily a special kind of language found in "Old Irish" texts and characteristic of conservative and even archaizing genres, namely law and poetry.⁹ It surely does preserve an earlier stage of the language, but arguably reflects only certain registers of that earlier language. It is thus possible that the article had developed even in the archaic period reflected by the Archaic Old Irish texts, but was excluded (perhaps by conscious convention) from these very tradition-bound genres.¹⁰ --- There is also another conceivable explanation. The Celtic article evolved from an old

⁹ On the problem of sorting out what is meant by the various senses of "Archaic Old Irish", see Greene 1977.

¹⁰ Parallel behavior in Welsh can be cited; Evans (1976:24) notes that the article is rare in early poetry, but common in the Old Welsh glosses and prose fragments.

demonstrative *sindos. Now there is at least one language in my sample, the Athapaskan language Slave, for which a Semitic-like placement restriction applies not to a full-fledged article but to an article-like demonstrative ("hedged article"). It is possible, accordingly, that a Semitic-like rule penetrated Insular Celtic very early, at a stage when *sindos was still a demonstrative (en route to article); the result would have been an early Irish situation like that found in Slave. This would sidestep the chronological difficulty.¹¹

The problems with Continental Celtic are at once chronological and geographical. First of all, such anomalies apparently are found only in Gaulish. Celtiberian, the most conservative of the Continental Celtic dialects, appears to show no CHS features whatever.¹² This in itself would seem geographically anomalous: Iberia, the Celtic region closest to Africa, is the least CHS-like. The Celtiberian corpus, however, comes almost entirely from inland sites, well away from the coast. And whatever North African influence there may have been on the prehistoric languages of Iberia, its locus most plausibly would not have been the inland regions but the coast. Similarly, migrants to Britain from North Africa might be expected to have made landfall in Spain en route northward, and perhaps even to have left linguistic traces --- but again the

¹¹ *sindos (or rather its reflex in) is attested in Archaic Old Irish; "but, even as a demonstrative, in is very rare in archaic texts" (Greene 1977:18).

¹² I would like to thank Joe Eska for insights and comments on this and a great many other points regarding Continental Celtic. Much of the discussion in this section is based on information he very generously provided to me, which I hope I have not misinterpreted.

probable locus would be near the sea. The noncoastal geographical distribution of Celtiberian, accordingly, makes its non-CHS-like nature less troublesome.

In Gaulish, several CHS features are clearly in evidence:

- [3] RC1 linker: Rel particle, not Rel pronoun
- [5] Special relative verb form
- [6] Polypersonal verb
- [7] Infixing/suffixing alternation

In more detail, the IE relative pronoun (preserved as such in Celtiberian) has become an invariant particle **-yo* [3], which occurs suffixed to the verb and thus might be taken as engendering a "special relative form" [5] (see discussion in sec. 3.2 [5b], also 5.2 [5f]). And features [6] and [7] have jointly to do with the grammaticalization of second-position clitics in Celtic (see sec. 3.2 [6b], and discussion in sec. 1.5.4). Polypersonal verbs are not, however, the only way of coding pronominal objects: nonclitic (tonic) object pronouns exist as well.

The Gaulish anomalies are both less and more of a problem than those involving stages of Insular Celtic. As remarked, they do not in themselves interfere with the workability of a substratal scenario in the British Isles. Their effect, rather, is to weaken the plausibility of such an account, by showing that CHS features manifestly could arise in Celtic at an early date even without substratal action.

The most facile way to dispose of these objections would be to posit substratal action on the continent as well. This is a pis aller; none of the special considerations applicable in the case of Insular

Celtic --- the large number of features, their exotic nature, the unusual "cut" of Insular Celtic --- applies to Gaulish. One might also propose to view the Gaulish CHS phenomena in areal terms, as fringe effects of the CHS language type radiating southeasterly from Britain to the Continent; the source of such a radial spread might be early Insular Celtic or even the pre-Celtic substratal language prior to the arrival of the Celts in Britain. However, geography argues against such a proposal, for the continental CHS phenomena are attested largely at the opposite extreme of Gaulish territory: northern Italy, southern and central France.

Can any other explanation be offered? Feature [3], the one skewed-inclusive item in the feature set, is probably the worst candidate for purely substratal explanation even in Insular Celtic; given the strong typological markedness of the IE-style relative pronoun, a language (whether Gaulish or early Insular Celtic) might well spontaneously alter its linker type. Regarding feature [5], the analysis depends on the degree to which we can consider **yo* and its host verb to be truly univerbated. This and the remaining features involve clitics and their degree of fusion with their host word --- analytically tricky matters in any context (recall the discussion in sec. 4.3). The infixing/suffixing alternation is reminiscent of the situation in Old Lithuanian (sec. 3.2 [6c]), where, however, univerbation is not at issue. In general, given the very small corpus of Gaulish that is ever likely to be available to us, it is hard to judge to what extent particular attestations of "Verb+Clitic" can or should be taken as manifesting a firmly grammaticized univerbation; there is little evidence for a clear and strong phonological bond of the sort seen in Irish.

The suggestions just given have a quality of special pleading. The above phenomena, however they are analyzed or explained away, do indeed seem to have arisen in Gaulish from IE forebears with no substratal help; they may not have developed with the full richness of their Insular Celtic counterparts, but they clearly represent very similar or even identical phenomena. Here it is important to emphasize a point made repeatedly in this study (secs. 1.5.4, 2.3.1.2, 7.3), that genetic filiation and substratal action are not mutually incompatible. What was nascent in Gaulish underwent much fuller development in Insular Celtic (see Eska 1993), where a substratum could have contributed to the process. The presence of these features in Gaulish does weaken somewhat the evidential force of the substratal argument; but it must be stressed that 16 of the 20 features --- and, significantly, all the exotics except one (feature [7]) --- occur only in Insular Celtic.

The last diachronic anomaly I will mention has a totally different geographical context. In discussing the development of word-initial change in Berber (sec. 3.2 [16d]), I remarked that the Berber phenomenon is unlikely to be ancient, probably not predating the Arab conquest. How, then, can it be relevant to the CHS problem? From a Berber perspective, a different question comes to the fore: where and why did Berber acquire the principle of initial change? Certainly not from Semitic or Egyptian, which have no trace of the phenomenon. On the other hand, initial change seems to have a natural home in West Africa, as mentioned in the areal discussion in sec. 6.4. Berber, therefore, is likely to have acquired initial change as an African areal feature. If so, there must have existed even in ancient times a language with

initial change that was spoken roughly in present-day Berber territory. Not Berber, but this language --- perhaps Afroasiatic, perhaps not --- would make a likely candidate for the ancient pre-Celtic substratal language, and thus a plausible source for the principle of initial mutation in Celtic.¹³

7.5 Assessment

The CHS problem has been worked over for so many years that it is important, by way of conclusion, to consider how the present study differs from its predecessors, and what it has to contribute both methodologically and factually. Its most fundamental contribution is not that it "solves" the particular problem at hand, but that it lays out and puts into practice, for the first time, a nonimpressionistic typological method for evaluating the significance of purely structural resemblances among unrelated languages. This method is not specific to the CHS problem. Whenever any set of structural similarities is observed between two language groups, the phenomena can be examined in global terms. The similarities can always be recast as realizations of higher-level features, and these higher features will have crosslinguistic distributions which can be determined straightforwardly by empirical sampling. In any concrete case, these distributions may or may not show

¹³ Though Celtic initial mutation originated from a metanalysis of older IE grammatical endings, the question of why the process happened at all, and happened only in Celtic, is not addressed by an IE-internal account. Initial mutation is not a very common phenomenon crosslinguistically, and it is legitimate to ask why a language should have come up with the concept of mutations as a phonosyntactic technique.

a preponderance of exotic features; they may or may not show striking areal restrictions; they may or may not reveal interesting correlations among features. The evidential value of the resemblances will depend on the observed distributions. If most of the features have a flat distribution, it will be relatively difficult to counter the charge that the resemblances are just coincidence; if most are exotics, coincidence becomes far less likely. If the features show strong correlations among themselves, the distribution probably conceals the germ of a multiparameter structural type. If there is little in the way of correlation, or if the observed correlations are strongly linked with a particular geographical macroarea, then some kind of historical explanation becomes relatively likelier. If the overall syndrome occurs nowhere else in the world, or if it tends to occur chiefly in some particular geographical region, then again the likelihood of historical explanation rises. In brief, every step involved in the analysis of the CHS problem can be applied to other cases in the same way; the outcome, of course, may differ.

One can find fault with the detailed implementation of the analysis as worked out in this particular study. The choice of universe dimension cannot avoid a degree of subjectivity; the scoring system used is imperfect, especially as regards the implicit zero treatment of blanks; the correlational procedure is very rough, in part reflecting the conceptual difficulties inherent in correlating privative and equipollent features. But none of these criticisms of detail should lessen the usefulness or validity of the method per se.

The typological method has important implications for the way

nongenetic problems are viewed by the historical linguistic community. It is precisely the existence of a solid nonimpressionistic method, the comparative method, that has imparted to genetic diachrony its special authority and status in modern historical linguistics; for only in this subdomain has there been a way to make reliable projections into prehistory even in the absence of concrete historical data, projections concerning the detailed articulation of a family's genetic evolution. The present method, similarly, should make it possible to treat a different class of problems in a nonimpressionistic way, enabling linguists to judge the relative likelihood of historical vs. ahistorical explanation in cases of purely structural resemblances between geographically remote, unrelated languages --- again in the absence of concrete historical data. There has heretofore been no method for treating problems of this sort. The method would seem readily applicable to other long-distance cases analogous to the CHS problem. One potential example is the remarkable crosslinguistic similarity of pidgins and creoles. The method might also be applied in evaluating proposals of deep genetic relatedness between distant languages, insofar as such proposals appeal to structural resemblances; here the typological method will enable us to address the logically prior question of whether such structural resemblances are likely to have a historical explanation at all.

In fact, the method can be useful even when the languages being compared are not geographically remote. In setting up linguistic areas, linguists are routinely faced with the question of whether the alleged areal feature is truly something special. Sometimes the feature exists in a contiguous zone extending beyond the putative region; sometimes it is sufficiently common worldwide to raise the possibility of

coincidence. The typological method, applied in exactly the same way as with the CHS problem, can help answer such questions. With it we can assess empirically the likelihood of coincidence, and determine the actual global distribution of the particular feature. Indeed, any concrete application of the method may have areal spin-offs quite unrelated to the original problem the method was applied to, for the assembled mass of crosslinguistic data may show unforeseen areal patternings. In this way the study of the CHS problem has pointed our attention to Africa, ultimately suggesting directions and providing material for areal analysis within a purely African context and having nothing to do with the CHS problem.

It could be objected that conclusions reached by the typological method cannot possibly attain to the level of reliability we have come to expect from the comparative method, and therefore cannot "prove" anything. This is perfectly true. However, one of the central themes running through this study is that the opposition of proof vs. speculation is a false dichotomy. Worse, it is not only false but harmful, for it tends to discourage any serious thinking about "unprovable" problems. The typological method is probabilistic, and as such represents a legitimate middle ground between proof and speculation.

The Celtic/Hamito-Semitic problem is undoubtedly the ideal testing ground for such a method, for here the set of similarity features is strikingly rich both quantitatively and qualitatively. Depending as it does on probabilistic argumentation, the method should work best where many features are involved. The typological method provides a nonimpressionistic response to almost all criticisms that have been

levelled against the CHS problem since its inception. The criteria for structural similarity are laid out explicitly and are tightly controlled. Hunches sprung from the methodological romanticism of a Pokorny or a Wagner can now be examined and tested, enabling us to distinguish between wishful thinking (a connection with Basque) and genuine insight (a West African link). The method confronts the issue of coincidence head-on by demonstrating for the first time the specialness of the CHS resemblances. Nearly half the features are exotics, making coincidence maximally unlikely as an explanation. Empirical observation, not hunch, shows that the full-fledged CHS type recurs nowhere else in the world. On the other hand, clear and objective structural parallels exist pointing specifically to northern and western Africa. The CHS type does involve some degree of inter-feature correlations, but much of that correlation can be laid to the CHS languages themselves and to Africa.

Increasingly over the course of this study, it is the exotics which have come to the fore as the key to the CHS problem. It is these features which are skewed-exclusive, which have little meaningful global correlation, which are maximally unmotivated vis-a-vis Indo-European, and which show considerable affinity for Africa; conversely, the nonexotic features tend to have a flat distribution, do show some degree of correlation, and to some extent show up with CHS values even in pre-Insular Celtic (Gaulish). This primacy of exotic features, it seems to me, is not accidental. Of course the typological method can be applied to any set of resemblances having any sort of global distribution. But in practice, a linguist would be unlikely to make the effort unless s/he had a strong preliminary intuition that the resemblances were in some

sense special. And exoticness is just the typological way of giving more exact expression to the notion of specialness.

This study also strongly underscores the worth of typology as a tool for uncovering aspects of prehistory which cannot be ferreted out in any other way.¹⁴ As such, it may help to counter recent trends downplaying the perceived value of typology in diachronic studies. In the last several decades typology has been appealed to uncritically by some scholars as a warrant for carrying out certain kinds of word order reconstruction¹⁵ --- an overenthusiasm which has tended to give typology a bad name among orthodox historical linguists.¹⁶ But to condemn the misuse of typology is not to condemn typology per se. Nichols argues that "typology has great potential for historical linguistics, particularly if historical linguistics has some say in the choice of typological parameters" (1986:101). The present study has explored this potential theoretically, methodologically, and concretely within the context of a particularly difficult problem, that of the CHS resemblances.

The one thing that the typological method cannot do is to articulate prehistoric scenarios. It has provided an "existence proof", so to speak (though "proof" is of course too strong a term) --- a demonstration that prehistoric contact of some sort fits the facts better than

¹⁴ In this respect it falls into the same tradition as Nichols 1986, 1992.

¹⁵ Thus e.g. Lehmann 1978a, 1978b, Vennemann 1974; see critique in Gensler 1984 (Chapter 3).

¹⁶ Witness Watkins's rejection of "a theory [of word order change] which elevates some of Greenberg's extremely interesting quasi-universals to the dubious status of an intellectual straitjacket" (1976:306).

the other explanatory alternatives. More specifically, the areal part of the analysis points strongly to some prehistoric link between the British Isles and northern or western Africa. Just what this link was is left open. A straightforward substratal scenario seems to me quite reasonable, and the ancient Barbary ape skull (sec. 1.5.3) might provide weak evidence for direct seafaring contact between the two regions in prehistoric times. But this is obviously a question for the prehistorian, not the linguist; and exploring the detailed mechanism of prehistoric contact was not the business of this purely linguistic study.¹⁷

The lack of a concrete prehistoric scenario, however, does not mean that there has been no advance in the status of the nonlinguistic problem. Prior to this study, a prehistorian would have had no basis beyond hunch or predilection for viewing the Celtic/Hamito-Semitic issue as something worth taking seriously. Now there is such a basis: a good linguistic case can be made for contact, couched not in impressionistic but in solid typological terms. The linguistic case is surely not proven, but it is about as good as it can be; contact, and specifically with Africa, is seen as the most likely reason for the set of resemblances. This could and should be an impetus to look at the issue afresh as a real problem in prehistory and archaeology --- despite the current archaeological trend to antidiffusionism. The problem is a

¹⁷ Conceivably the African language presumed to have been in contact with Britain was not Hamito-Semitic at all. This is not very likely; Berber and Egyptian score far higher than the various "second-tier" languages of Africa. Yet the diachronic evidence (sec. 6.5) does suggest that the CHS type may well have been more widespread and more firmly installed in Africa millennia ago than it is today.

valid one, and well worth the effort.¹⁸

We return at the end to the radical skeptic. It may seem good sober linguistic practice to say of a problem, this remains unproven and even unprovable. But to say that something is unproven and therefore to dismiss it out of hand is not good science. It has the result that fascinating problems are stigmatized and left unexamined. Previous approaches to the Celtic/Hamito-Semitic problem have indeed been impressionistic and overenthusiastic. This does not vitiate the substratalist approach per se, if applied cautiously; it is important not to confuse the critique of a problem with a critique of the methodology applied to that problem. This study has presented a new way of evaluating the likelihood of various modes of explanation for crosslinguistic resemblances of a certain kind. The case is not proven, and probably never can be; but taking a typological approach gives us a probabilistic middle way that makes even an unprovable problem tractable.

¹⁸ An analogy from genetic linguistics will make the point clearly. Purely on the basis of the linguistic evidence, the linguist can state with confidence that there is (say) a genetic link between Wiyot/Yurok and Algonquian. The prehistorian can then proceed with confidence to try to find such a link in the archaeological record. The present case involves a lower confidence level, but the principle is the same.

General bibliography

[Note: Works which served as primary data sources for the language survey are generally not listed here, but in a separate reference list which immediately follows this main bibliography. A few such works are referenced in passing in the main text.]

Abbreviations:

- BBCS = Bulletin of the Board of Celtic Studies
- BLS = Berkeley Linguistics Society (Proceedings, Annual Meeting)
- BSL = Bulletin de la Société de Linguistique de Paris
- CLS = Chicago Linguistic Society (Proceedings, Annual Meeting)
- CTIL = Current Trends in Linguistics
- EC = Etudes Celtiques
- GLECS = Groupe Linguistique d'Etudes Chamito-Sémitiques (Comptes Rendus)
- IF = Indogermanische Forschungen
- IJAL = International Journal of American Linguistics
- JIES = Journal of Indo-European Studies
- KZ = Zeitschrift für vergleichende Sprachforschung (Kuhns Zeitschrift)
- MSS = Münchener Studien zur Sprachwissenschaft
- NTS = Norsk Tidsskrift for Sprogvidenskap
- PBA = Proceedings of the British Academy
- PRIA = Proceedings of the Royal Irish Academy
- RC = Revue Celtique
- RIA = Dictionary of the Irish Language (Royal Irish Academy)
- SC = Studia Celtica

TPS = Transactions of the Philological Society
 UCPL = University of California Publications in Linguistics
 WPLU = Working Papers on Language Universals (Stanford)
 ZCP = Zeitschrift für Celtische Philologie
 ZDMG = Zeitschrift der deutschen morgenländischen Gesellschaft

- Adams, G. B. 1956. Irish toponymy and the theories of pre-Gaelic substrata. *Bulletin of the Ulster Place-Name Society* 4(1):1-23.
 [apparently unavailable in the United States]
- . 1970. Language and man in Ireland. *Ulster Folklife* 15/16:140-71.
- . 1975. Hamito-Semitic and the pre-Celtic substratum in Ireland and Britain. *Hamito-Semitica*, ed. by J. & T. Bynon, pp. 233-47. *Janua Linguarum, Series Practica*, no. 200. The Hague: Mouton.
- . 1980. Place-names from pre-Celtic languages in Ireland and Britain. *Nomina* 4:46-63.
- Ahlqvist, Anders. 1977. Typological notes on Irish word order. *Studies in Descriptive and Historical Linguistics: Festschrift for Winfred P. Lehmann*, ed. by Paul J. Hopper, pp. 267-81. Amsterdam: John Benjamins.
- . 1988. Remarks on the question of dialects in Old Irish. *Historical dialectology*, ed. by Jacek Fisiak, pp. 23-38. Berlin: Mouton de Gruyter.
- Ammerman, Albert J., and L. L. Cavalli-Sforza. 1984. The neolithic

transition and the genetics of populations in Europe. Princeton:
Princeton University Press.

Avienus, Rufus Festus. Ora Maritima. See Murphy, J. P.

Basset, André. 1949. Sur le participe berbère. GLECS 5:34-36.

---. 1952. La langue berbère. London: Oxford University Press.

Basset, René. 1886. Manuel de langue kabyle. Paris: Librairie orientale
et américaine.

Baudiš, Josef. 1913a. Nominalsätze und Nominalprädikate im Irischen. ZCP
9:309-35.

---. 1913b. Zum Gebrauch der Verbalnomina im Irischen. ZCP 9:380-417.

---. 1922. On the character of the Celtic languages (part 1). RC 39:33-
46.

---. 1923. On the character of the Celtic languages (part 2). RC
40:104-26.

---. 1926. Glossy k nauce o keltských pronominech osobních a k otázce
novo-keltského typu jazykového [Notes on the study of the Celtic
personal pronouns and on the question of the Neoceltic language
type; in Czech]. Mněma [Festschrift Zubatý], pp. 215-28. Prague.

Bednarczuk, Leszek. 1987. An attempt at typological characterization of
the Celtic languages. Lingua Posnaniensis 28:7-13.

Bell, Alan. 1978. Language samples. Universals of human language, ed. by
Joseph Greenberg, vol. 1, pp. 123-56. Stanford: Stanford University

Press.

Bender, Marvin L. 1969. Chance CVC correspondences in unrelated languages. *Language* 45:519-31.

Benveniste, Emile. 1974. Convergences typologiques. Problèmes de linguistique générale (II), pp. 103-12. Gallimard. [original: *L'Homme* 6.2:5-12 (1966)]

Bergin, Osborn. 1938a. On the syntax of the verb in Old Irish. *Ériu* 12:197-214.

---. 1938b. The native Irish grammarian. *PBA* 205-35.

---. 1946. *Ériu* and the Ablaut. *Ériu* 14:147-53.

Binchy, D. A. 1962. Patrick and his biographers. *Studia Hibernica* 2:7-173.

Birnbaum, Henrik. 1970. Problems of typological and genetic linguistics viewed in a generative framework. *Janua Linguarum, Series Minor*, no. 106. The Hague: Mouton.

Blansitt, Edward L., Jr. 1975. Progressive aspect. *WPLU* 18:1-34.

Blok, H. P. 1948. Notes on localism in African languages. *Orientalia Neerlandica*, pp. 75-110. Leiden: A. W. Sijthoff.

Boas, Franz. 1917. Introductory. *IJAL* 1:1-8.

Boling, Bruce D. 1972. Some problems of the phonology and morphology of the Old Irish verb. *Ériu* 23:73-101.

- Borer, Hagit, and Laurice Tuller. 1985. Nominative/agreement complementarity and VSO order in Standard Arabic. *Precis from the 15th Conference on African Linguistics*, ed. by Russell G. Schuh, pp. 27-32. *Studies in African Linguistics*, Supplement 9.
- Bouda, Charles. 1949. Un substrat basque en celtique? *Eusko-Jakintza* 3:336-38.
- Brugmann, Karl. 1904. Die Demonstrativpronomina der indogermanischen Sprachen. (Abhandlungen der Königl. Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Philologisch-historische Klasse, 22.6). Leipzig: Teubner.
- . 1917. Der Ursprung des Scheinsubjekts 'es' in den germanischen und den romanischen Sprachen (Berichte der Königl. Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Philologisch-historische Klasse, 69.5). Leipzig: Teubner.
- Brugnatelli, Vermondo. 1984. Confronti tipologici e storici tra costrutti genitivali camito-semitici e indeuropei. *Studi semitici* (n.s.) 1:42-64.
- Byrne, Francis John. 1962. Review of Teach Yourself Irish, by Myles Dillon and Donncha Ó Cróinín. *ZCP* 29:198-99.
- Byrne, James. 1885. General principles of the structure of language. Two vols. London: Trübner & Co.
- Calder, George. 1917. *Auraicept na n-Éces: The scholars' primer*. Edinburgh: John Grant.

- Camaj, Martin. 1984. Albanian grammar. Wiesbaden: Otto Harrassowitz.
- Campanile, Enrico. 1976. Indo-European and non-Indo-European elements in the Celtic dialects. *JIES* 4:131-38.
- Černý, Jaroslav, and Sarah Israelit Groll. 1984. A Late Egyptian grammar. Rome: Biblical Institute Press.
- Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. *Subject and topic*, ed. by Charles N. Li, pp. 25-55. New York: Academic Press.
- Chaker, Salem. 1983. Un parler berbère d'Algérie (Kabylie): Syntaxe. Aix en Provence: Université de Provence Service des Publications.
- Christian Brothers, The. 1980. New Irish grammar. Dublin: C. J. Fallon.
- Comrie, Bernard. 1989. Language universals and linguistic typology (2nd edition). Chicago: University of Chicago Press.
- Comrie, Bernard, and Norval Smith. 1977. *Lingua descriptive studies: Questionnaire*. *Lingua* 42:1-71.
- Covington, Michael A. 1984. Syntactic theory in the High Middle Ages. Cambridge: Cambridge University Press.
- Cowgill, Warren. 1975. The origins of the Insular Celtic conjunct and absolute verbal endings. *Flexion und Wortbildung*, ed. by Helmut Rix, pp. 40-70. Wiesbaden: Ludwig Reichert.
- . 1987. The distribution of infixed and suffixed pronouns in Old Irish. *Cambridge Medieval Celtic Studies* 13:1-5.

- Craddock, Jerry R. 1969. Latin legacy versus substratum residue. UCPL 53. Berkeley and Los Angeles: University of California Press.
- Daniel, Glyn. 1963. The megalith builders of Western Europe. Baltimore: Penguin Books.
- Davies, John. 1621. *Antiquae linguae britannicae, nunc communiter dictae Cambro-Britannicae, a suis Cymraecae vel Cambricae, ab alijs Wallicae, Rudimenta...* London (apud Iohannem Billium). [Reprinted 1968: *English linguistics*, 1500-1800, no. 70; Scolar Press] [Welsh grammar]
- . 1632. *Antiquae linguae britannicae, nunc vulgo dictae Cambro-britannicae, a suis Cymraecae vel Cambricae, ab alijs Wallicae, et linguae latinae, dictionarium duplex.* London (in aedibus R. Young). [Reprinted 1968: *English linguistics*, 1500-1800, no. 99; Scolar Press] [Welsh-Latin dictionary]
- Delbrück, Berthold. 1888. *Altindische Syntax.* Halle: Buchhandlung des Waisenhauses.
- Denez, Per. 1972. *Brezhoneg... buan hag aes.* Paris: Editions Omnivox.
- Desbordes, Yann. 1983. *Petite grammaire du breton moderne.* Mouladurioù hor yezh.
- Dictionary of the Irish Language (compact edition). 1983. Dublin: Royal Irish Academy.
- Dillmann, August. 1907. *Ethiopic grammar.* Trans. by James A. Crichton. London: Williams & Norgate.

- Dillon, Myles. 1926-27. Nominal predicates in Irish. ZCP 16:313-56.
- . 1927-28. Nominal predicates in Irish (continued). ZCP 17:307-46.
- . 1943. On the structure of the Celtic verb. Language 19:252-55.
- . 1945. Linguistic borrowing and historical evidence. Language 21:12-17.
- . 1947. Celtic and the other Indo-European languages. TPS 15-24.
- . 1955. On the syntax of the Irish verb. TPS 104-16.
- Disterheft, Dorothy. 1980. The syntactic development of the infinitive in Indo-European. Columbus, Ohio: Slavica.
- Dixon, R. M. W. 1979. Ergativity. Language 55:59-138.
- Dryer, Matthew S. 1989. Large linguistic areas and language sampling. Studies in Language 13:257-92.
- Edwards, Charles. [1677] 1936. Hanes y ffydd ddiffuant (ed. G. J. Williams). Cardiff.
- Eska, Joseph F. 1993. Rethinking the evolution of Celtic constituent configuration. ms.
- Evans, D. Ellis. 1967. Gaulish personal names. Oxford: Clarendon Press.
- . 1983. Language contact in pre-Roman and Roman Britain. Aufstieg und Niedergang der römischen Welt, ed. by Wolfgang Haase, 29.2:949-87.
- . 1986. The Celts in Britain (up to the formation of the Brittonic languages): History, culture, linguistic remains, substrata.

Geschichte und Kultur der Kelten, ed. by Karl Horst Schmidt and Rolf Ködderitzsch, pp. 102-15. Heidelberg: Carl Winter.

---. 1991. Celticity, identity and the study of language --- fact, speculation and legend. *Archaeologia Cambrensis* 140:1-16.

Evans, D. Simon. 1950. Similarities between the Welsh and Hebrew languages. *National Message*, pp. 85a-85g.

---. 1971. Concord in Middle Welsh. *SC* 6:42-56.

---. 1976. A grammar of Middle Welsh. Dublin: Institute for Advanced Studies.

Evans, Emrys. 1989/90. Heinrich Wagner [necrology]. *SC* 24/25:157-60.

Falc'hun, F. 1976. L'article déterminé en Gaulois. *Word* 28:37-50.

Feehan, Fanny. 1981. Suggested links between Eastern and Celtic music. *The Celtic consciousness*, ed. by Robert O'Driscoll, pp. 333-39. New York: George Braziller.

Fife, James. 1990. The semantics of the Welsh verb: A cognitive approach. Cardiff: University of Wales Press.

Finck, Franz Nikolaus. 1899. Der deutsche Sprachbau als Ausdruck deutscher Weltanschauung. Marburg: N. G. Elwert'sche Verlagsbuchhandlung.

---. 1901. Die Klassifikation der Sprachen. Marburg: N. G. Elwert'sche Verlagsbuchhandlung.

- . 1907. Der angebliche passivische Charakter des transitiven Verbs.
KZ 41:209-82.
- . 1909. Die Haupttypen des Sprachbaus. Leipzig: B. G. Teubner.
- Fleuriot, Léon. 1964. Le vieux Breton: Eléments d'une grammaire. Paris:
Klincksieck.
- Fokos-Fuchs, D. R. 1962. Rolle der Syntax in der Frage nach Sprach-
chverwandtschaft. Wiesbaden: Otto Harrassowitz.
- Fraenkel, E. 1928. Review of Mnēma [cf. Baudiš 1926]. Slavia 7:359-67.
- Fraser, J. 1912. A use of the verbal noun in Irish. Miscellany presented
to Kuno Meyer, ed. by Osborn Bergin and Carl Marstrander, pp. 216-
26. Halle: Max Niemeyer.
- Friedrich, Johannes. 1974. Hethitisches Elementarbuch (I). Heidelberg:
Carl Winter.
- Gagnepain, Jean. 1961. A propos du "verbe celtique" [review article on
Wagner 1959]. EC 9:309-26.
- Gaidoz, H. 1917. Deux érudits gallois: John Rhys et Llywarch Reynolds.
Revue Internationale de l'Enseignement 71:12-21, 104-14, 193-213,
274-85, 375-82.
- Gardiner, Alan. 1957. Egyptian grammar. London: Oxford University Press.
- Gary, Judith Olmsted, and Saad Gamal-Eldin. 1982. Cairene Egyptian col-
loquial Arabic. Amsterdam: North-Holland.

- Gensler, Orin. 1984. What is typology good for anyway? University of California, Berkeley. ms.
- . 1989. From Proto-Semitic to Geez: The mechanism of word order shift in Ethiopic Semitic. Paper presented at the December meeting of the Linguistic Society of America (1989-90).
- . 1991. The syntactic conservatism of Arabic within Semitic. Paper presented at the March meeting of the North American Conference on Afroasiatic Linguistics (NACAL).
- . 1992. Head/dependent marking and tightness of syntactic bond: The Welsh-Berber paradox. Paper presented at the January meeting of the Linguistic Society of America (1991-92).
- Gesenius, W. 1910. Gesenius' Hebrew grammar. Revised by E. Kautzsch; English translation by A. E. Cowley. Oxford: Clarendon Press.
- Gordon, Cyrus H. 1965. Ugaritic textbook. *Analecta Orientalia* 38. Rome: Pontificium Institutum Biblicum.
- Greenberg, Joseph. 1966a. Some universals of language with particular reference to the order of meaningful elements. *Universals of language*, ed. by Joseph Greenberg, pp. 73-113. Cambridge, Ma.: MIT Press.
- . 1966b. *The languages of Africa*. Bloomington: Indiana University Press.
- . 1978. How does a language acquire gender markers?. *Universals of human language*, ed. by Joseph H. Greenberg et al., 3:47-82.

Stanford: Stanford University Press.

---. 1980. Circumfixes and typological change. Papers from the 4th International Conference on Historical Linguistics, ed. by Elizabeth C. Traugott et al., 233-241. Amsterdam: Benjamins.

---. 1987. Language in the Americas. Stanford: Stanford University Press.

Greene, David. 1966a. The making of Insular Celtic. Proceedings of the Second International Congress of Celtic Studies (1963), pp. 123-36. Cardiff: University of Wales Press.

---. 1966b. The Irish language (An Ghaeilge). Cork: Mercier Press.

---. 1969. *Varia* II. *Ériu* 21:89-98.

---. 1977. Archaic Irish. *Indogermanisch und Keltisch*, ed. by Karl Horst Schmidt, pp. 11-33. Wiesbaden: Ludwig Reichert.

---. 1983. The coming of the Celts: The linguistic viewpoint. Proceedings of the Sixth International Congress of Celtic Studies (1979), ed. by Gearoid Mac Eoin, pp. 131-37. Dublin: Institute for Advanced Studies.

Gumperz, John J., and Robert Wilson. 1971. Convergence and creolization: A case from the Indo-Aryan/Dravidian border. *Pidginization and creolization of languages*, ed. by Dell Hymes, pp. 151-67. Cambridge: Cambridge University Press.

Gutenbrunner, Siegfried. 1961/62. Review of Wagner 1959. *Archiv für das Studium der neueren Sprachen* 198:172-73.

- Haarmann, Harald. 1976. Aspekte der Arealtypologie: Die Problematik der europäischen Sprachbunde. Tübingen: TBL Verlag Gunter Narr.
- Hahn, E. Adelaide. 1969. Naming-constructions in some Indo-European languages. American Philological Association monograph 27. Case Western Reserve University Press.
- Hamp, Eric P. 1973. Inordinate clauses in Celtic. You take the high node and I'll take the low node (paravolume to CLS 9), pp. 229-51.
- Hanoteau, A. 1896. Essai de grammaire de la langue tamachek'. Algiers: Adolphe Jourdan.
- Harbison, Peter. 1975. The coming of the Indo-Europeans to Ireland: An archaeological viewpoint. JIES 3:101-19.
- Harden, Donald. 1963. The Phoenicians. London: Thames & Hudson.
- Hartmann, Hans. 1954. Das Passiv: Eine Studie zur Geistesgeschichte der Kelten, Italiker und Arier. Heidelberg: Carl Winter.
- . 1960. Der Typus ocus é im Irischen. Indogermanica: Festschrift für Wolfgang Krause, pp. 8-23. Heidelberg: Carl Winter.
- Havers, Wilhelm. 1931. Handbuch der erklärenden Syntax. Heidelberg: Carl Winter.
- Hawkes, C. F. C. 1975. Pytheas: Europe and the Greek explorers. Oxford.
- Hawkins, John A. 1983. Word order universals. New York: Academic Press.
- Heine, Bernd, Ulrike Claudi, and Friederike Hünemeyer. 1991. From cognition to grammar: Evidence from African languages. Approaches to

- grammaticalization, ed. by Elizabeth Traugott and Bernd Heine, Vol. 1, pp. 149-87. Amsterdam: John Benjamins.
- Heine, Bernd, and Mechthild Reh. 1984. Grammaticalization and reanalysis in African languages. Hamburg: Helmut Buske.
- Hemon, Roparz. 1984. A historical morphology and syntax of Breton. Dublin: Institute for Advanced Studies.
- Herren, Michael. 1979. Some new light on the life of Virgilius Maro Grammaticus. *PRIA* 79C, no. 2:27-71.
- Heusler, Andreas. 1950. *Altisländisches Elementarbuch*. Heidelberg: Carl Winter.
- Hewitt, Steve. 1984-85. Quelques ressemblances structurales entre le breton et l'arabe: conséquence d'une typologie ordinale commune? *La Bretagne linguistique* 1:223-62.
- Hillgarth, J. N. 1984. Ireland and Spain in the seventh century. *Peritia* 3:1-16.
- Hodge, Carleton T. 1970. The linguistic cycle. *Language Sciences* 13:1-7 (December 1970).
- Holland, Gary B. 1980. Problems of word order change in selected Indo-European languages. PhD dissertation, University of California, Berkeley.
- . 1982. A note on raising in Indo-European. *BLS* 8:158-69.
- Holmer, Nils M. 1969. A Celtic-Hittite correspondence. *Ériu* 21:23-24.

- Holmer, Nils M., Ants Uesson, and Olof Smedberg. 1961. On linguistic types, blood groups and culture areas [with maps]. *Språkliga Bidrag* 4.16:5-40.
- Hook, Peter E. 1983. South Asia as a linguistic area: discriminating the grain of history from the chaff of universals. *Festschrift for Henry Hoenigswald*, ed. by George Cardona and N. Zide, pp. 155-68. Tübingen: Gunter Narr.
- Hubschmid, Johannes. 1953. *Sardische Studien*. Bern: A. Francke.
- . 1960. *Mediterrane Substrate*. Bern: A. Francke.
- Huemer, Johann. 1886. *Virgilii Maronis Grammatici Opera*. Leipzig: B. G. Teubner.
- Hymes, Dell. 1956. Na-Déné and positional analysis of categories. *American Anthropologist* 58:624-38.
- Jackson, Kenneth. 1953. *Language and history in early Britain*. Edinburgh: Edinburgh University Press.
- . 1955. The Pictish language. *The problem of the Picts*, ed. by F. T. Wainwright, pp. 129-66. Edinburgh: Thomas Nelson & Sons.
- . 1962. Review of Wagner 1959. *Studia Hibernica* 2:229-34.
- James, Allan. 1987. *John Morris-Jones*. Cardiff: University of Wales Press. [biography]
- Jeffers, Robert J. 1976. Typological shift and change in complex sentence structure. *CLS Parasession on Diachronic Syntax*, pp. 136-49.

- Jespersen, Otto. 1922. *Language*. London: George Allen & Unwin.
- Johnson, H. H. 1904/05. The heresy of connecting Welsh and Semitic, etc. [sic]. *Celtic Review* 1:160-72.
- Jones, G. Lloyd. 1983. *The discovery of Hebrew in Tudor England: A third language*. Manchester: Manchester University Press.
- Jones, John Morris. 1891. *Y Cymry [The Welsh]. Y Gwyddoniadur Cymreig [Encyclopaedia cambrensis]* 3:95-103.
- . 1900. Pre-Aryan syntax in Insular Celtic. In Rhys and Brynmor-Jones, Appendix B, pp. 617-41.
- . 1904/05. Professor Johnson's heresy hunt. *Celtic Review* 1:276-81.
- . 1913. *A Welsh grammar*. Oxford: Clarendon Press.
- . 1931. *Welsh syntax: An unfinished draft*. Cardiff: University of Wales Press.
- [---. 1986. John Morris-Jones: *Llyfryddiaeth anodiadol* [annotated bibliography], compiled by Huw Walters. Aberystwyth: National Library of Wales.]
- Jongeling, K. 1988(?). *Welsh and Hebrew, some typological remarks*. pp. 1-18. ms.
- . 1991. On the VSO character of Hebrew. *Studies in Hebrew and Aramaic syntax*, ed. by K. Jongeling et al., pp. 103-11. Leiden: E. J. Brill.

- Joseph, Brian D. 1983. The synchrony and diachrony of the Balkan infinitive. Cambridge: Cambridge University Press.
- Kendrick, T. D. 1950. *British antiquity*. London: Methuen & Co.
- Koch, John T. 1983. The sentence in Gaulish. *Proceedings of the Harvard Celtic Colloquium* 3:169-215.
- . 1985. Movement and emphasis in the Gaulish sentence. *BBCS* 32:1-37.
- . 1991. Ériu, Alba, and Letha: When was a language ancestral to Gaelic first spoken in Ireland? *Emania* [Bulletin of the Navan Research Group, Belfast] 9:17-27.
- Krámský, Jiří. 1972. The article and the concept of definiteness in language. *Janua Linguarum, Series Minor*, no. 125. The Hague: Mouton.
- Lambdin, Thomas O. 1978. *Introduction to Classical Ethiopic (Ge'ez)*. Missoula, Montana: Scholars Press.
- Lansberg, William Ross. 1940. Evidence of a Mediterranean substratum in Celtic. *Language* 16:95-98.
- Lehmann, Winfred P. 1978a. The great underlying ground-plans. *Syntactic typology*, ed. by Winfred P. Lehmann, pp. 3-55. Austin: University of Texas Press.
- . 1978b. Conclusion: Toward an understanding of the profound unity underlying languages. *Syntactic typology*, ed. by Winfred P. Lehmann, pp. 395-432. Austin: University of Texas Press.

- Lejeune, Michel. 1980. Notes d'étymologie gauloise: V. Les démonstratifs. EC 17:51-54.
- Lejeune, Michel, et al. 1985. Textes gaulois et gallo-romains en cursive latine: 3. Le plomb du Larzac. EC 22:95-177.
- Leslau, Wolf. 1945. The influence of Cushitic on the Semitic languages of Ethiopia: A problem of substratum. Word 1:59-82.
- Lewis, Henry. 1942. The sentence in Welsh. PBA 259-80.
- . 1956. Old Welsh syntax. Celtica 3:295-99.
- Lewis, Henry, and Holger Pedersen. 1961. A concise comparative Celtic grammar. Göttingen: Vandenhoeck & Ruprecht.
- Lewy, Ernst. 1942. Der Bau der europäischen Sprachen. PRIA 48C, no. 2:15-117.
- . 1949. The old linguist's beliefs. Reprinted in in *Kleine Schriften*, pp. 129-38.
- . 1951. Die Lehre von den Sprachtypen. Reprinted in *Kleine Schriften*, pp. 9-21.
- . 1961. *Kleine Schriften*. Berlin: Akademie-Verlag.
- . 1967. Versuch einer kurzen Beschreibung des Altirischen [posthumous]. *Beiträge zur Indogermanistik und Keltologie: Julius Pokorny zum 80. Geburtstag*, ed. by Wolfgang Meid, pp. 217-22. Innsbruck: Innsbrucker Beiträge zur Kulturwissenschaft 13.

- Lieber, Rochelle. 1987. An integrated theory of autosegmental processes. Albany, NY: SUNY Press. [Chapter 2: Mutation]
- Locker, Ernst. 1961. Die Sprachschichten Nordostafrikas. Anzeiger der Österreichischen Akademie der Wissenschaften, Philosophisch-Historische Klasse 98(18):140-53.
- . 1962. Die ältesten Sprachschichten Westeuropas. Sitzungsberichte der Österreichischen Akademie der Wissenschaften, Philosophisch-Historische Klasse 240(5).
- Longacre, Robert E. 1990. Storyline concerns and word order typology in East and West Africa. Studies in African Linguistics, Supplement 10.
- Lundman, Bertil. 1957. Einige neue Arbeiten zur Anthropologie Irlands [with maps]. Homo [Göttingen] 8:168-73.
- Lynch, John. 1975. Oral/nasal alternation and the realis/irrealis distinction in Oceanic languages. Oceanic Linguistics 14:87-99.
- Macalister, R. A. Stewart. 1938-39. Lebor Gabála Éirenn [The book of the taking of Ireland] (parts 1, 2). Dublin: Irish Texts Society.
- Mac Cana, Proinsias. 1973. On Celtic word-order and the Welsh 'abnormal' sentence. Ériu 24:90-120.
- McCone, Kim. 1979. Pretonic preverbs and the absolute verbal ending in Old Irish. Ériu 30:1-34.
- . 1985. The Würzburg and Milan glosses: Our earliest sources of 'Middle Irish'. Ériu 36:85-106.

- Mackie, Euan. 1977. *The megalith builders*. Oxford: Phaidon.
- McManus, Damian. 1991. *A guide to ogam*. Maynooth: An Sagart.
- Mac Mathúna, Seamus. 1989. Heinrich Wagner [necrology]. *EC* 26:215-17.
- McNally, Robert E. 1958. The "tres linguae sacrae" in early Irish Bible exegesis. *Theological Studies* 19:395-403.
- Mac White, Eóin. 1955. Problems of Irish archaeology and Celtic philology. *ZCP* 25:1-29.
- Mallory, J. P. 1989. *In search of the Indo-Europeans*. London: Thames and Hudson.
- Marcy, G. 1936. Note sur le pronom relatif-sujet et le pseudo-participe dans les parlers berbères. *BSL* 37.1:45-57.
- . 1939. Fonctions originales dans les parlers berbères des pronoms démonstratifs-relatifs *id, *in. *BSL* 40.2:151-73.
- Maro, Vergilius. See Huemer.
- Martinet, Andre. 1952. Celtic lenition and Western Romance consonants. *Language* 28:192-217.
- Masica, Colin P. 1976. *Defining a linguistic area: South Asia*. Chicago: University of Chicago Press.
- Matisoff, James A. 1992. *The mother of all morphemes: Augmentatives and diminutives in areal and universal perspective*. ms.
- Meid, Wolfgang. 1963. *Die indogermanischen Grundlagen der altirischen*

- absoluten und konjunkten Verbalflexion. Wiesbaden: Otto Harrassowitz.
- . 1967 (ed.). Beiträge zur Indogermanistik und Keltologie: Julius Pokorny zum 80. Geburtstag gewidmet. Innsbrucker Beiträge zur Kulturwissenschaft 13.
- . 1972. Old Celtic languages. CTIL 9(2):1190-1201. The Hague: Mouton.
- . 1986. The Celtic languages. Geschichte und Kultur der Kelten, ed. by Karl Horst Schmidt and Rolf Ködderitzsch, pp. 116-22. Heidelberg: Carl Winter.
- Meillet, Antoine. 1964. Introduction à l'étude comparative des langues indoeuropéennes (8th edition). University, Ala.: University of Alabama Press.
- . 1965[1912]. L'évolution des formes grammaticales. Linguistique historique et linguistique générale, vol. 1, pp. 130-48. Paris: Librairie Honoré Champion.
- Meinhof, Carl. 1912. Die Sprachen der Hamiten. Abhandlungen des Hamburgischen Kolonialinstituts 9. Hamburg: L. Friederichsen & Co.
- Meyer, Kuno. 1901. The present state of Celtic studies. Celtica 1:167-69.
- . 1913a. Aus dem Nachlass Heinrich Zimmers. ZCP 9:87-120.
- . 1913b. Learning in Ireland in the fifth century and the transmission of letters. Dublin: School of Irish Learning.
- Misteli, Franz. 1893. Charakteristik der hauptsächlichsten Typen des

Sprachbaues (part 2 of Abriss der Sprachwissenschaft, by H. Steinthal and Franz Misteli). Berlin: Ferd. Dümmlers Verlagsbuchhandlung.

Moravcsik, Edith A. 1969. Determination. WPLU 1:64-98.

Morgan, Prys. 1973/74. Boxhorn, Leibniz, and the Welsh. SC 8/9:220-28.

Morris-Jones, John. See Jones, John Morris.

Moscatti, Sabatino. 1968. The world of the Phoenicians. Trans. by Alastair Hamilton. London: Weidenfeld and Nicolson.

Mourant, A. E., Ada C. Kopeć, and Kazimiera Domaniewska-Sobczak. 1958. The ABO blood groups: Comprehensive tables and maps of world distribution. Oxford: Blackwell Scientific Publications.

---. 1976. The distribution of the human blood groups. 2nd edition. London: Oxford University Press.

Mourant, A. E., and I. Morgan Watkin. 1952. Blood groups, anthropology and language in Wales and the Western countries. Heredity 6:13-36.

Murphy, J. P. 1977. Rufus Festus Avienus: Ora maritima, or description of the seacoast. Chicago: Ares Publishers.

Myhill, John M. 1984. A study of aspect, word order, and voice. PhD dissertation, University of Pennsylvania.

Nichols, Johanna. 1973. The Balto-Slavic predicate instrumental: A problem in diachronic syntax. PhD dissertation, University of California, Berkeley.

- . 1986. Head-marking and dependent-marking grammar. *Language* 62:56-119.
- . 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Ó Cléirigh, C. R. 1959. Review of Wagner 1959. *Béaloides* 27:124-25.
- O'Donovan, John. 1845. *A grammar of the Irish language*. Dublin: Hodges & Smith.
- Oftedal, Magne. 1968. North European geography of sounds. *Scottish Gaelic Studies* 11:248-58. [review of Wagner 1964]
- Ora Maritima. See Murphy, J. P.
- Oswalt, Robert L. 1991. A method for assessing distant linguistic relationships. Sprung from some common source: *Investigations into the prehistory of languages*, ed. by Sydney M. Lamb and E. Douglas Mitchell, pp. 389-404. Stanford: Stanford University Press.
- Parry, Thomas. 1972. *John Morris-Jones*. Cardiff: University of Wales Press. [biography]
- Penhallurick, R. D. 1986. *Tin in antiquity*. London: The Institute of Metals.
- Pezron, Paul. 1706. *The antiquities of nations; More particularly of the Celtae or Gauls*. London. [Translation by D. Jones of Pezron's Antiquités de la nation et de la langue des Celtes (1703)]
- Piggott, Stuart. 1983. *The coming of the Celts: The archaeological*

argument. *Proceedings of the Sixth International Congress of Celtic Studies* (1979), ed. by Gearoid Mac Eoin, pp. 138-48. Dublin: Institute for Advanced Studies.

- Pokorny, Julius. 1926. Das nicht-indogermanische Substrat im Irischen (Parts 1,2). *ZCP* 16:95-144, 231-66.
- . 1927. Das nicht-indogermanische Substrat im Irischen (Part 3). *ZCP* 16:363-94.
- . 1928a. Das nicht-indogermanische Substrat im Irischen (Part 4). *ZCP* 17:373-88.
- . 1929. Das nicht-indogermanische Substrat im Irischen (Part 5). *ZCP* 18:233-48.
- . 1928b. Die Theorie der Substrate und die Entstehung des Indogermanischen. *Actes du premier congrès des linguistes* [10-15 April 1928], pp. 175-76. Leiden: A. W. Sijthoff.
- . 1936. Substrattheorie und Urheimat der Indogermanen. *Mitteilungen der anthropologischen Gesellschaft in Wien* 66:69-91.
- . 1949. Zum nichtindogermanischen Substrat im Inselkeltischen. *Die Sprache* 1:235-45.
- . 1951. Sprachliche Beziehungen zwischen dem alten Orient und den britischen Inseln. *Archiv Orientální* 19:268-70.
- . 1953. Keltologie [annotated bibliographical essay]. *Allgemeine und vergleichende Sprachwissenschaft*, by Vittore Pisani [Indo-European] and Julius Pokorny [Celtic], pp. 95-186, 194-99. Bern: A. Francke.

- . 1955. Zur irischen Nomenbildung und Urgeschichte. MSS 7:56-67.
- . 1957. Gedanken zur Indogermanenfrage. *Mnēmēs charin* (Gedenkschrift Paul Kretschmer), ed. by Heinz Kronasser, 2:83-85. Vienna: Verlag der Wiener Sprachgesellschaft.
- . 1959. Keltische Urgeschichte und Sprachwissenschaft. *Die Sprache* 5:152-64.
- . 1960a. The pre-Celtic inhabitants of Ireland. *Celtica* 5:229-40.
- . 1960b. Review of Wagner 1959. *ZCP* 28:137-41.
- . 1962. Die Sprache der vorkeltischen Bewohner Nordwesteuropas. II. Fachtagung für indogermanische und allgemeine Sprachwissenschaft (1961). *Innsbrucker Beiträge zur Kulturwissenschaft, Sonderheft* 15:129-38. [German reworking of 1960a]
- . 1964. Zur Anfangsstellung des inselkeltischen Verbuns. MSS 16:75-80.
- [---. 1967. Bibliographie der wissenschaftlichen Veröffentlichungen von Julius Pokorny, compiled by Hans Schmeja. *Beiträge zur Indogermanistik und Keltologie: Julius Pokorny zum 80. Geburtstag*, ed. by Wolfgang Meid, pp. 323-32. Innsbruck: *Innsbrucker Beiträge zur Kulturwissenschaft* 13.]
- Porzig, Walter. 1942. *Die Namen für Satzinhalte im Griechischen und im Indogermanischen*. Berlin: Walter de Gruyter.
- Powell, T. G. E. 1958. *The Celts*. New York: Praeger.

- Preusler, Walther. 1956. Keltischer Einfluss im Englischen. *Revue des langues vivantes* 22:322-50.
- Quinn, Bob. 1986. *Atlantean*. London: Quartet Books.
- Renfrew, Colin. 1979. *Before civilization: The radiocarbon revolution and prehistoric Europe*. Cambridge: Cambridge University Press.
- (ed.). 1983. *The megalithic monuments of Western Europe*. London: Thames & Hudson.
- . 1987. *Archaeology and language: The puzzle of Indo-European origins*. London: Jonathan Cape.
- Renouf, P. le Page. 1875. *An elementary grammar of the ancient Egyptian language*. London: Samuel Bagster & Sons.
- Rhys, John. 1890. Traces of a non-Aryan element in the Celtic family. *Scottish Review* 16:30-47.
- Rhys, John, and David Brynmor-Jones. 1900. *The Welsh people*. London: T. Fisher Unwin Ltd.
- Ridgeway, William. 1907. Who were the Romans? *PBA* 17-60.
- Robert, Gruffydd. 1567. *Gramadeg cymraeg*. [Reprinted 1939, London: Oxford University Press] [Welsh grammar]
- Rohlf, Gerhard. 1930. Vorlateinische Einflüsse in den Mundarten des heutigen Italiens? *Germanisch-Romanische Monatsschrift* 18:37-56.
- Rosén, Hannah. 1981. *Studies in the syntax of the verbal noun in early Latin*. Munich: Wilhelm Fink Verlag.

- Rössler, Otto. 1950. Verbalbau und Verbalflexion in den Semito-
hamitischen Sprachen. ZDMG 100:461-514.
- . 1958. Die Sprache Numidiens. Sybaris (Festschrift Hans Krahe), pp.
94-120. Wiesbaden: Otto Harrassowitz.
- . 1960. Review of Wagner 1959. ZCP 28:141-47.
- Rowlands, Henry. 1723. *Mona Antiqua Restaurata*. Dublin. [reprint 1979,
NY: Garland]
- Sadiqi, Fatima. 1986. *Studies in Berber syntax*. Würzburg: Königshausen &
Neumann.
- Salesbury [Salisbury], Henry. 1593. *Grammatica Britannica*. [Reprinted
1969: English linguistics, 1500-1800, no. 189; Scolar Press] [Welsh
grammar]
- Salvaneschi, Enrica. 1978. Latino biblico quale tramite di concordanze
celtico-semitiche? *Lingue a contatto nel mondo antico*, pp. 61-92.
Pisa: Giardini.
- Sammes, Aylett. 1676. *Britannia antiqua illustrata: or, The antiquities
of ancient Britain, derived from the Phoenicians*.
- Sandfeld, Kr. 1930. *Linguistique balkanique: Problèmes et résultats*.
Paris: Klincksieck.
- Sapir, J. David. 1971. West Atlantic: An inventory of the languages,
their noun class systems and consonant alternation. CTIL 7:45-112.
- Schaller, Helmut. 1975. *Die Balkansprachen*. Heidelberg: Carl Winter.

- Schmid, Wolfgang P. 1960. Review of Wagner 1959. IF 65:312-16.
- Schmidt, Karl Horst. 1963. Review of Wagner 1959. ZDMG 113:347-51.
- . 1976. Der Beitrag der keltiberischen Inschrift von Botorrita zur Rekonstruktion der protokeltischen Syntax. Word 28:51-62.
- . 1982. Keltisch-Hethitisches. Serta Indogermanica: Festschrift für Günter Neumann zum 60. Geburtstag, ed. by Johann Tischler, pp. 357-62. Innsbrucker Beiträge zur Sprachwissenschaft 40.
- . 1989. Heinrich Wagner [necrology]. ZCP 43:226-29.
- . 1990. The postulated pre-Indo-European substrates in Insular Celtic and Tocharian. When worlds collide: Indo-Europeans and Pre-Indo-Europeans [Linguistica Extranea, Studia 19], ed. by T. L. Markey and John A. C. Greppin, pp. 179-202. Ann Arbor: Karoma.
- Schmoll, Ulrich. 1959. Die Sprachen der vorkeltischen Indogermanen Hispaniens und das Keltiberische. Wiesbaden: Otto Harrassowitz.
- Segert, Stanislav. 1976. A grammar of Phoenician and Punic. Munich: C. H. Beck.
- Senn, Alfred. 1966. Handbuch der litauischen Sprache (I). Heidelberg: Carl Winter.
- Sjoestedt-Jonval, M.-L. 1939. Review of Pokorny, Zur Urgeschichte der Kelten und Illyrier. BSL 40.3:105-111.
- Smalley, Beryl. 1952. The study of the Bible in the Middle Ages. Oxford: Basil Blackwell.

- Sommerfelt, Alf. 1960. Mixed languages versus remodelled languages. NTS 19:316-26.
- . 1962. Review of Wagner 1959. Lochlann 2 (NTS suppl. vol. 6): 229-31.
- Spitzer, Leo. 1928. Bericht über den ersten internationalen Linguistenkongress im Haag (10.-15. April 1928). Die Neueren Sprachen 36:439-44.
- Stern, Ludwig Christian. 1901. Review of Rhys, Celtic Folklore. ZCP 3:605-13.
- Strachan, John. 1909. An introduction to early Welsh. Manchester: University of Manchester Press.
- Stumme, Hans. 1899. Handbuch des Schilhischen von Tazerwalt. Leipzig: J. C. Hinrichs.
- Talmy, Leonard. 1978. Relations between subordination and coordination. Universals of human language, ed. by Joseph Greenberg, vol. 4, pp. 487-513. Stanford: Stanford University Press.
- Tardi, D. 1928. Les Epitomae de Virgile de Toulouse: Essai de traduction critique. Paris: Boivin et Cie.
- Thiel, Matthias. 1973. Grundlagen und Gestalt der Hebräischkenntnisse des frühen Mittelalters. Biblioteca degli "Studi Medievali", 4. Spoleto, Italy.
- Thomas, Charles. 1976. Imported late-Roman Mediterranean pottery in Ireland and western Britain: Chronologies and implications. PRIA 76C,

no. 13:245-55.

Thomason, Sarah Grey, and Terrence Kaufman. 1988. Language contact, creolization, and genetic linguistics. Berkeley & Los Angeles: University of California Press.

Thurneysen, Rudolf. 1930. Lat. · = Ir. edón. ZCP 18:427-28.

---. 1946. A grammar of Old Irish. Trans. by D. A. Binchy and Osborn Bergin. Dublin: Institute for Advanced Studies.

Tovar, Antonio. 1971. Konvergenz, Kontakt, Zufall: Irische Beispiele. Interlinguistica: Sprachvergleich und Übersetzung, ed. by Karl-Richard Bausch & Hans-Martin Gauger, pp. 96-100. Tübingen: Max Niemeyer.

Trubetzkoy, N. S. 1939. Gedanken über das Indogermanenproblem. Acta Linguistica 1:81-89.

Ultan, Russell. 1970. Some sources of consonant gradation. WPLU 2:C1-C30.

Uspensky, B. 1968. Principles of structural typology. Janua Linguarum, Series Minor, no. 62. The Hague: Mouton.

Vallancey, Charles. 1786. An essay on the antiquity of the Irish language. Collectanea de rebus hibernicis, 2:8 (pp. 251-336). Dublin: Luke White.

Vendryes, Joseph. 1911. Ludwig-Christian Stern [necrology]. RC 32:518-19.

- . 1929. Sir John Morris Jones [necrology]. RC 46:427-29.
- . 1937. La position linguistique du celtique. PBA 333-71.
- . 1956. Sur l'emploi impersonnel du verbe. Celtica 3:185-97.
- Vennemann, Theo. 1974. Topics, subjects, and word order: From SXV to SVX via TVX. Historical linguistics, ed. by John M. Anderson and Charles Jones, 1:339-76. Amsterdam: North-Holland.
- Von Soden, Wolfram. 1969. Grundriss der akkadischen Grammatik. Rome: Pontificium Institutum Biblicum.
- Vycichl, Werner. 1957. L'article défini du berbère. Mémorial André Basset, pp. 139-46. Paris: Adrien Maisonneuve.
- Wagner, Heinrich. 1951. Die sprachgeographische Stellung des Germanischen und der germanischen Sprachen. Utrecht: H. J. Smits.
- . 1959. Das Verbum in den Sprachen der britischen Inseln. Tübingen: Niemeyer.
- . 1964. Nordeuropäische Lautgeographie. ZCP 29:225-98.
- . 1967. Zur unregelmässigen Wortstellung in der altirischen Alliterationsdichtung. Beiträge zur Indogermanistik und Keltologie: Julius Pokorny zum 80. Geburtstag, ed. by Wolfgang Meid, pp. 289-314. Innsbruck: Innsbrucker Beiträge zur Kulturwissenschaft 13.
- . 1969. The origin of the Celts in the light of linguistic geography. TPS 1969: 203-50.

- . 1970a. Studies in the origins of early Celtic civilisation. ZCP 31:1-58.
- . 1970b. Ernst Lewy [necrology]. ZCP 31:288-91.
- . 1972a. Beiträge in Erinnerung an Julius Pokorny. ZCP 32:1-89.
- . 1972b. Julius Pokorny [necrology]. ZCP 32:313-19.
- . 1976. Common problems concerning the early languages of the British Isles and the Iberian peninsula. Actas del I Coloquio sobre Lenguas y Culturas Prerromanas de la Peninsula Iberica, ed. by Francisco Jordá et al., pp. 387-407. Acta Salamanticensia, Filosofia y Letras, 95, Universidad de Salamanca.
- . 1977. Wortstellung im Keltischen und Indogermanischen. Indogermanisch und Keltisch, ed. by Karl Horst Schmidt, pp. 204-35. Wiesbaden: Ludwig Reichert.
- . 1978. The typological background of the ergative construction. PRIA 78C, no. 3:37-74.
- . 1981. Near Eastern and African connections with the Celtic world. The Celtic consciousness, ed. by Robert O'Driscoll, pp. 51-67. New York: George Braziller.
- . 1985. Das Hethitische vom Standpunkte der typologischen Sprachgeographie. Testi Linguistici 7. Pisa: Giardini.
- . 1987. The Celtic invasions of Ireland and Great Britain: Facts and theories. ZCP 42:1-40.

- Wainwright, F. T. (ed.). 1955. The problem of the Picts. Edinburgh: Thomas Nelson & Sons.
- Walters, Huw. 1986. John Morris-Jones: Llyfryddiaeth Anodiadol [annotated bibliography]. Aberystwyth: National Library of Wales.
- Watkins, Calvert. 1962. Indo-European origins of the Celtic verb: I. The sigmatic aorist. Dublin: Institute for Advanced Studies.
- . 1963. Preliminaries to a historical and comparative analysis of the syntax of the Old Irish verb. *Celtica* 6:1-49.
- . 1976. Towards Proto-Indo-European syntax: Problems and pseudo-problems. *CLS Diachronic Syntax* volume, pp. 305-26.
- Watkins, T. Arwyn. 1957. Yr arddodiad HG. (h)i, in; CC. y (=yn), yn. *BBCS* 17:137-58.
- . 1960. CC. y/yn berfenwol. *BBCS* 18:362-72.
- . 1962. Ffuriant a chystrawen y geirynnau adferfol/traethiadol mewn Cymraeg, Cernyweg a Llydaweg. *BBCS* 19:295-315.
- Weinreich, Uriel. 1963. Languages in contact. The Hague: Mouton.
- Welmers, William E. 1973. African language structures. Berkeley and Los Angeles: University of California Press.
- Williams, G. J. 1973/74. The history of Welsh scholarship. *SC* 8/9:195-219.
- Williams, Stephen J. 1980. A Welsh grammar. Cardiff: University of Wales Press.

Wolff, Philippe. 1971. *Western languages, AD 100-1500*. New York:
McGraw-Hill.

Wright, W. 1967. *A grammar of the Arabic language*. 3rd edition. Cambridge: Cambridge University Press.

Zimmer, Heinrich. 1910. Über direkte Handelsverbindungen Westgalliens mit Irland im Altertum und frühen Mittelalter. *Sitzungsberichte der königlich preussischen Akademie der Wissenschaften*, pp. 1031-1119.

---. 1913. Aus dem Nachlass Heinrich Zimmers. See Meyer, Kuno.

Zimmer, Stefan. 1986. *Briefe deutscher Keltologen an John Morris-Jones*. ZCP 41:280-86.

Bibliography to Appendix 2

[Note: This bibliography includes works which served as primary data sources for the language survey in Appendix 2.]

Abbreviations:

BLS = Proceedings of the Berkeley Linguistics Society

CWH = Chicago Which Hunt: see s.v. Peranteau

UCPL = University of California Publications in Linguistics

Abraham, R. C. 1958. Dictionary of Modern Yoruba. London: University of London Press.

---. 1962. Dictionary of the Hausa language. London: University of London Press.

Ackerman, Farrell. 1987. Miscreant morphemes: Phrasal predicates in Ugric. PhD dissertation, University of California, Berkeley.

Ahenakew, Freda. 1987. Cree language structures: A Cree approach. Winnipeg: Pemmican Publications.

Amin-Madani, Sadegh, and Dorothea Lutz. 1972. Persische Grammatik. Heidelberg: Julius Groos Verlag.

- Applegate, Richard B. 1972. Ineseño Chumash grammar. PhD dissertation, University of California, Berkeley.
- . 1974. Ineseño Chumash dictionary. ms.
- Armbruster, Charles Hubert. 1960. Dongolese Nubian: A grammar. Cambridge: Cambridge University Press.
- . 1965. Dongolese Nubian: A lexicon. Cambridge: Cambridge University Press.
- Aronson, Howard I. 1972. Some notes on relative clauses in Georgian. CWH 136-43.
- . 1982. Georgian: A reading grammar. Columbus, Ohio: Slavica.
- Aryanpur-Kashani, Abbas and Manoochehr. 1978. The concise Persian-English dictionary. Tehran: Amir Kabir.
- Asher, R. E. 1985. Tamil. London: Croom Helm.
- Awobuluyi, Oladele. 1978. Essentials of Yoruba grammar. Ibadan: Oxford University Press.
- Bamgboṣe, Ayo. 1966. A grammar of Yoruba. Cambridge: Cambridge University Press.
- . 1974. A short Yoruba grammar. Ibadan: Heinemann Educational Books.
- Bliese, Loren F. 1981. A generative grammar of Afar. Dallas: Summer Institute of Linguistics.
- Bradley, C. Henry, and Barbara E. Hollenbach, eds. 1988 (vol. 1), 1990

- (vol. 2). Studies in the syntax of Mixtecan languages. Dallas: Summer Institute of Linguistics. [sketches of 5 Mixtec dialects]
- Callaghan, Catherine A. 1963. A grammar of the Lake Miwok language. PhD dissertation, University of California, Berkeley.
- . 1965. Lake Miwok dictionary. UCPL 39.
- Camaj, Martin. 1984. Albanian grammar, trans. Leonard Fox. Wiesbaden: Otto Harrassowitz.
- Campbell, Lyle. 1985. The Pipil language of El Salvador. Mouton Grammar Library, 1. Berlin: Mouton.
- Capell, Arthur. 1941. A new Fijian dictionary. Sydney: Australian Medical Publishing Co., Ltd.
- Cohen, Marcel. 1936. *Traité de langue amharique*. Paris: Institut d'Ethnologie.
- Cole, Peter. 1982. Imbabura Quechua. Amsterdam: North-Holland.
- Comrie, Bernard. 1981. The languages of the Soviet Union. Cambridge: Cambridge University Press. [pp. 266-72, 276-77 on Gilyak/Nivkh]
- Cowan, J. Ronayne, and Russell G. Schuh. 1976. Spoken Hausa. Ithaca: Spoken Language Services.
- Creissels, Denis. 1983. *Eléments de grammaire de la langue mandinka*. Grenoble: Publications de l'Université des Langues et Lettres.
- Creissels, Denis, Sidia Jatta, and Kalifa Jobarteh. 1982. *Lexique mandinka-français*. Mandenkan 3 (whole volume).

- Davies, John. 1989 [1981]. *Kobon*. London: Routledge.
- Dayley, Jon P. 1985. *Tzutujil grammar*. UCPL 107.
- . 1989a. *Tümpisa (Panamint) Shoshone grammar*. UCPL 115.
- . 1989b. *Tümpisa (Panamint) Shoshone dictionary*. UCPL 116.
- Derbyshire, Desmond C. 1979. *Hixkaryana*. Amsterdam: North-Holland.
- . 1985. *Hixkaryana and linguistic typology*. Dallas: Summer Institute of Linguistics.
- De Rijk, Rudolf P. G. 1972. *Relative clauses in Basque: A guided tour*. CWH 115-35.
- Dixon, R. M. W. 1972. *The Dyirbal language of North Queensland*. Cambridge: Cambridge University Press.
- . 1988. *A grammar of Boumaa Fijian*. Chicago: University of Chicago Press.
- Dover, Kenneth J. 1960. *Greek word order*. Cambridge: Cambridge University Press.
- Elbert, Samuel H., and Mary Kawena Pukui. 1979. *Hawaiian grammar*. Honolulu: University Press of Hawaii.
- English, Leo James. 1986. *Tagalog-English dictionary*.
- Fal, Arame, Rosine Santos, and Jean Léonce Doneux. 1990. *Dictionnaire wolof-français*. Paris: Editions Karthala.

- Falkenstein, Adam. 1959. Das Sumerische. Handbuch der Orientalistik I:2:[1-2]:1. Leiden: E. J. Brill.
- Foley, William A. 1991. The Yimas language of New Guinea. Stanford: Stanford University Press.
- Fortescue, Michael. 1984. West Greenlandic. London: Croom Helm.
- Friedrich, Johannes. 1952. Hethitisches Wörterbuch. Heidelberg: Carl Winter.
- . 1974. Hethitisches Elementarbuch, 1. Teil. Heidelberg: Carl Winter.
- Gragg, Gene B. 1972. [Relative clauses in] Sumerian and selected Afro-Asiatic languages. CWH 153-68.
- Gordon, Lynn. 1986. Maricopa morphology and syntax. UCPL 108.
- Hagman, Roy S. 1977. Nama Hottentot grammar. Bloomington: Indiana University Publications.
- Haiman, John. 1980. Hua: A Papuan language of the Eastern Highlands of New Guinea. Amsterdam: John Benjamins.
- Hartmann, Josef. 1980. Amharische Grammatik. Wiesbaden: Franz Steiner.
- Hasegawa, Yoko. 1992. Syntax, semantics, and pragmatics of TE-linkage in Japanese. PhD dissertation, University of California, Berkeley.
- Hawkins, Emily A. 1979. Hawaiian sentence structures. Pacific Linguistics, Series B, no. 61.
- . 1982. Pedagogical grammar of Hawaiian. Honolulu: University Press

of Hawaii.

Hewitt, B. G. 1979. Abkhaz. Amsterdam: North-Holland.

Hills, Robert A. 1990. A syntactic sketch of Ayutla Mixtec. In Bradley and Hollenbach 1990, pp. 1-260.

Hinds, John. 1986. Japanese. London: Croom Helm.

Howeidy, A. 1953. Concise Hausa grammar. London: George Ronald.

Jakobson, Roman. 1971. Notes on Gilyak. Selected Writings II:72-102. The Hague: Mouton.

Kimball, Geoffrey D. 1991. Koasati grammar. Lincoln, Nebraska: University of Nebraska Press.

Kraft, Charles H. 1963. A study of Hausa syntax. 3 volumes. Hartford Studies in Linguistics 9. Hartford: Hartford Seminary Foundation.

Kraft, Charles H., and A. H. M. Kirk-Greene. 1973. Hausa. New York: David McKay.

Kreinovich, E. A. 1934. Nivkhskij (giljatskij) jazyk. Jazyki i pis'mennost' narodov severa, Part 3: Jazyki i pis'mennost' paleoaziatskikh narodov, pp. 181-222. Moscow: Gosudarstvennoe uchebno-pedagogicheskoe izdatel'stvo.

Kuipers, Aert H. 1967. The Squamish language. Janua Linguarum, Series Practica 73 (also vol. 2, 1969). The Hague: Mouton.

Lafitte, Pierre. 1962. Grammaire basque. Bayonne: Editions des "Amis du Musée Basque" et "Ikas".

- Lambdin, Thomas O. 1983. Introduction to Sahidic Coptic. Macon, Ga.: Mercer University Press.
- Lambton, Ann K. S. 1974. Persian grammar. Cambridge: Cambridge University Press.
- Langdon, Margaret. 1977. Semantics and syntax of expressive "say" constructions in Yuman. BLS 3:1-11.
- Lehmann, Thomas. 1989. A grammar of modern Tamil. Pondicherry: Pondicherry Institute of Linguistics and Culture.
- Leslau, Wolf. 1967. Amharic textbook. Wiesbaden: Harrassowitz.
- . 1976. Concise Amharic dictionary. Berkeley & Los Angeles: University of California Press.
- Lewis, G. L. 1967. Turkish grammar. Oxford: Oxford University Press.
- Li, Charles N., and Sandra A. Thompson. 1981. Mandarin Chinese: A functional reference grammar. Berkeley: University of California Press.
- Lorimer, D. L. R. 1935. The Burushaski language. Volume 1: Introduction and grammar. Instituttet for Sammenlignende Kulturforskning, Serie B: Skrifter 29, 1. Oslo.
- . 1938. The Burushaski language. Volume 3: Vocabularies and index. Instituttet for Sammenlignende Kulturforskning, Serie B: Skrifter 29, 3. Oslo.
- Lotz, Janos. 1939. Das ungarische Sprachsystem. Stockholm: Ungarisches Institut.

- Luraghi, Silvia. 1990. Old Hittite sentence structure. London: Routledge.
- MacDonald, Lorna. 1990. A grammar of Tauya. Mouton Grammar Library, 6. Berlin: Mouton de Gruyter.
- Mann, Stuart E. 1948. A historical Albanian-English dictionary. London: William Clowes & Sons.
- Matisoff, James A. 1973. The grammar of Lahu. UCPL 75.
- . 1988. The dictionary of Lahu. UCPL 111.
- Matsumoto, Yoshiko. 1988. Grammar and semantics of adnominal clauses in Japanese. PhD dissertation, University of California, Berkeley.
- Merlan, Francesca. 1982. Mangarayi. Amsterdam: North-Holland.
- Morgan, J. L. 1972. Some aspects of relative clauses in English and Albanian. CWH 63-72.
- Newmark, Leonard, Philip Hubbard, and Peter Prifti. 1982. Standard Albanian: A reference grammar for students. Stanford: Stanford University Press.
- Nichols, Johanna. 1990. Ingush: A grammatical sketch. Ms; to appear in: Indigenous languages of the Caucasus, Vol. 3, ed. J. Greppin et al.
- Njie, Codu Mbassy. 1982. Description syntaxique du wolof de Gambie. Dakar: Les Nouvelles Editions Africaines.
- Noonan, Michael. 1992. A grammar of Lango. Mouton Grammar Library, 7. Berlin: Mouton de Gruyter.

- Panfilov, V. Z. 1962-1965. *Grammatika nivkhsского jazyka* (vol. I, 1962; vol. II, 1965). Moscow & Leningrad: Nauk.
- Payne, Doris L. 1986. Basic constituent order in Yagua clauses: Implications for word order universals. In *Handbook of Amazonian languages*, ed. Desmond C. Derbyshire and Geoffrey K. Pullum, 1:440-65. Berlin: Mouton de Gruyter.
- . 1990. The pragmatics of word order: Typological dimensions of verb initial languages. Berlin: Mouton de Gruyter.
- Payne, Doris L., and Thomas E. Payne. 1990. Yagua. In *Handbook of Amazonian languages*, ed. Desmond C. Derbyshire and Geoffrey K. Pullum, 2:249-474. Berlin: Mouton de Gruyter.
- Peranteau, Paul M., et al. 1972. *The Chicago Which Hunt: Papers from the Relative Clause Festival*. Chicago: Chicago Linguistic Society.
- Pukui, Mary Kawena, and Samuel H. Elbert. 1971. *Hawaiian dictionary*. Honolulu: University of Hawaii Press.
- Rebuschi, Georges. 1984. *Structure de l'énoncé en basque*. Paris: SELAF.
- Rice, Keren. 1989. *A grammar of Slave*. Mouton Grammar Library, 5. Berlin: Mouton de Gruyter.
- Roberts, John R. 1987. *Amele*. London: Croom Helm.
- Rowlands, E. C. 1969. *Teach yourself Yoruba*. London: English Universities Press.
- Saltarelli, Mario. 1988. *Basque*. London: Croom Helm.

- Samarin, William J. 1966. The Gbeya language. UCPL 44.
- Santos, Vito C. 1978. Pilipino-English dictionary.
- Schachter, Paul, and Fe T. Otanes. 1972. Tagalog reference grammar.
Berkeley: University of California Press.
- Schaub, Willi. 1985. Babungo. London: Croom Helm.
- Schütz, Albert J. 1985. The Fijian language. Honolulu: University Press
of Hawaii.
- Smyth, Herbert Weir. 1956. Greek grammar. Cambridge: Harvard University
Press.
- Taylor, Charles. 1985. Nkore-Kiga. London: Croom Helm.
- Thomas, David D. 1971. Chrau grammar. Oceanic Linguistics, Special Pub-
lication 7. Honolulu: University of Hawaii Press.
- Thomsen, Marie-Louise. 1984. The Sumerian language. Copenhagen: Akadem-
isk Forlag.
- Tschenkeli, Kita. 1958. Einführung in die georgische Sprache (Vol. 1).
Zurich: Amirani Verlag.
- . 1960-1974. Georgisch-Deutsches Wörterbuch. 3 volumes. Zurich:
Amirani Verlag.
- Tucker, A. N., and J. Tompo Ole Mpaayei. 1955. A Maasai grammar. London:
Longmans, Green & Co.
- Vogt, Hans. 1971. Grammaire de la langue géorgienne. Instituttet for

Sammenlignende Kulturforskning, Serie B: Skrifter 57, Oslo.

Watkins, Laurel J. 1984. A grammar of Kiowa. Lincoln: University of Nebraska Press.

Wolfart, H. Christoph. 1973. Plains Cree: A grammatical study. Philadelphia: American Philosophical Society.

Wolfart, H. Christoph, and Janet F. Carroll. 1981. Meet Cree: A guide to the language (2nd ed.). Lincoln: University of Nebraska Press.

Appendix 1A: Languages and sources
 ("[D]" indicates "Dictionary")
 For sources, see Bibliography to Appendix 2

Abkhaz: Hewitt 1979
 Afar: Bliese 1981
 Akkadian: Von Soden 1969
 Albanian: Camaj 1984, Newmark et al. 1982, Morgan 1972; Mann 1948 [D]
 Amele: Roberts 1987
 Amharic: Leslau 1967, Cohen 1936, Hartmann 1980; Leslau 1976 [D]
 Arabic (Classical): Wright 1967
 Arabic (Modern Cairene): Gary & Gamal-Eldin 1982
 Babungo: Schaub 1985
 Basque: Saltarelli 1988; Lafitte 1962, Rebuschi 1984
 Berber: Chaker 1983
 Burushaski: Lorimer 1935; Lorimer 1938 [D]
 Chinese (Mandarin): Li & Thompson 1981
 Chrau: Thomas 1971
 Chumash (Ineseno): Applegate 1972
 Coptic: Lambdin 1983
 Cree: Wolfart 1973; Wolfart & Carroll 1981, Ahenakew 1987
 Dyirbal: Dixon 1972
 Egyptian: Gardiner 1957
 English: personal knowledge
 Eskimo (West Greenlandic): Fortescue 1984
 Fijian (Boumaa): Dixon 1988; Capell 1941 [D], Schütz 1985
 French (modern spoken): personal knowledge
 Gbeya: Samarin 1966
 Geez (Classical Ethiopic): Dillmann 1907, Lambdin 1978
 Georgian: Aronson 1982, 1972, Vogt 1971, Tschenkeli 1958, 1960-1974 [D]
 Gilyak/Nivkh: Jakobson 1971, Comrie 1981, Panfilov 1962-1965,
 Kreinovich 1934
 Greek (classical): Smyth 1956
 Hausa: Kraft & Kirk-Greene 1973, Cowan & Schuh 1976, Kraft 1963,
 Howeidy 1953; Abraham 1962 [D]
 Hawaiian: Elbert & Pukui 1979; Hawkins 1979, 1982; Pukui & Elbert 1971 [D]
 Hebrew (Biblical): Gesenius 1910
 Hittite: Friedrich 1974, Luraghi 1990
 Hixkaryana: Derbyshire 1979; cf. Derbyshire 1985
 Hottentot (Nama): Hagman 1977
 Hua: Haiman 1980
 Hungarian: Lotz 1939; Ackerman 1987
 Ingush: Nichols 1990
 Irish (Old): Thurneysen 1946
 Irish (Modern): Christian Brothers 1980
 Japanese: Hinds 1986; Matsumoto 1988
 Kiowa: Watkins 1984
 Koasati: Kimball 1991
 Kobon: Davies 1989

Lahu: Matisoff 1973; 1988 [D]
 Lango: Noonan 1992
 Maasai: Tucker & Mpaayei 1955
 Mandinka: Creissels 1983; Creissels et al. 1982 [D]
 Mangarayi: Merlan 1982
 Maricopa: Gordon 1986
 Miwok (Lake): Callaghan 1963; 1965 [D]
 Mixtec (Ayutla): Hills 1990; other dialects in Bradley & Hollenbach
 1988-1990
 Nkore-Kiga: Taylor 1985
 Nubian (Dongolese): Armbruster 1960; 1965 [D]
 Persian (modern): Lambton 1974; Amin-Madani & Lutz 1972;
 Aryanpur-Kashani & Aryanpur-Kashani 1978 [D]
 Pipil: Campbell 1985
 Quechua (Imbabura): Cole 1982
 Shoshone (Panamint): Dayley 1989a; 1989b [D]
 Slave: Rice 1989
 Squamish: Kuipers 1967
 Sumerian: Thomsen 1984; Falkenstein 1959, Gragg 1972
 Tagalog: Schachter & Otañes 1972; English 1986 [D], Santos 1978 [D]
 Tamil: Lehmann 1989; Asher 1985
 Tauya: MacDonald 1990
 Turkish: Lewis 1967
 Tzutujil: Dayley 1985
 Welsh (Middle): Evans 1976
 Wolof: Njie 1982; Fal et al. 1990 [D]
 Yagua: Payne & Payne 1990; Payne 1986
 Yimas: Foley 1991
 Yoruba: Rowlands 1969, Awobuluyi 1978, Bamgbose 1966, 1974;
 Abraham 1958 [D]

Appendix 1B: Sample by areas and genetic groupings
 (* denotes CHS languages, including "supplementary" languages)

Africa:

Niger-Congo

Babungo (Grassfields Bantu)

Gbeya (Adamawa-Eastern)

Mandinka (Mande)

Nkore-Kiga (Bantu)

Wolof (West Atlantic)

Yoruba (Kwa)

Chari-Nile

Lango (Western Nilotic)

Maasai (Nilo-Hamitic = Eastern Nilotic)

Nubian (Eastern Sudanic)

Khoisan

Hottentot

Afroasiatic

Afar (Cushitic)

Hausa (Chadic)

* Berber

* Geez (Ethiopic Semitic); Amharic [later stage of Ethiopic]

* Egyptian; Coptic [later stage]

Near East and Caucasus:

Abkhaz (Northwest Caucasian)

Georgian (Kartvelian = South Caucasian)

Ingush (Northeast Caucasian)

Sumerian (isolate)

Turkish (Altaic)

Indo-European

Hittite (Anatolian)

Modern Persian (Iranian)

Semitic (Afroasiatic)

* Akkadian

* Classical Arabic; Modern Cairene Arabic [later stage]

* Biblical Hebrew

Europe (and Western Eurasia) (excluding Near East/Caucasus):

Basque (isolate)

Hungarian (Finno-Ugric)

Indo-European

Albanian

English (Germanic)

French (Romance)

Classical Greek

* Old Irish (Celtic); Modern Irish [later stage]

* Middle Welsh (Celtic)

North America:

Chumash (Hokan)
 Cree (Algonquian)
 Eskimo (Eskimo-Aleut)
 Kiowa (Kiowa-Tanoan)
 Koasati (Muskogean)
 Maricopa (Yuman [Hokan])
 Lake Miwok (Utian [Penutian])
 Shoshone (Uto-Aztecan)
 Slave (Athapaskan)
 Squamish (Salish)

Meso-America and South America:

Meso-America

Mixtec (Otomanguean)
 Pipil (Uto-Aztecan)
 Tzutujil (Mayan)

South America

Hixkaryana (Carib)
 Quechua (Andean)
 Yagua (isolate)

New Guinea and Australia:

Australia

Dyirbal (Pama-Nyungan)
 Mangarayi (non-Pama-Nyungan)

New Guinea

Amele (Gum [Mabuso])
 Hua (East-Central Highlands)
 Kobon (Kalam)
 Tauya (Brahman)
 Yimas (Lower Sepik)

Oceania:

Fijian (Oceanic = Eastern Austronesian)
 Hawaiian (Polynesian [Oceanic])
 Tagalog (Northwest Austronesian)

Eastern Eurasia:

Burushaski (isolate)
 Mandarin Chinese (Sino-Tibetan)
 Chrau (Mon-Khmer [Austroasiatic])
 Gilyak (isolate)
 Japanese (isolate)
 Lahu (Lolo-Burmese [Sino-Tibetan])
 Tamil (Dravidian)

Appendix 2: The language data

This appendix presents, in condensed form, relevant data from the grammars of all the non-CHS languages used in the survey, and also for two recent "neo-CHS" languages, Coptic and Amharic, which have diverged radically from their older forms. (For basic data on Celtic and Hamito-Semitic languages, see Chapter 3.) Page-number references into the grammar books are given as well. The intention is to provide a handy summary of the raw material used in scoring each language, and a reference key for those who wish to look into any particular question in greater depth. Each data summary follows the feature-by-feature format employed in Chapters 3 and 5; additionally, a preliminary section 0 appears, giving basic categorial information about nouns, verbs, and articles.

Little attempt has been made to reanalyze the information from the various sources so as to make it terminologically consistent. With rare exceptions --- notably where the description, analysis, or terminology in the source seemed wrong or obfuscatory, or where I had to form my own generalizations based on example sentences or texts --- the material is presented more or less as in the grammar. Lowercase labels will usually refer to grammatical bound morphemes, especially case endings, whereas labels beginning with a capital letter will usually refer either to categories per se or to free forms participating in a syntactic

construction; however, this schema may not have been carried through with total consistency. These distinctions should be borne in mind especially for the N-Gen construction, where (e.g.)

gen = genitive case marker poss = pronominal possessor affix

Gen = genitive (possessor) noun HeadN = head (possessee) noun

yielding notations like

Gen-gen HeadN-poss = "of-the-man his-house" = the man's house

Phonetic transcriptions are more or less as in the original source, except that O and E replace open-o and epsilon for typographical convenience.

Abkhaz (Caucasus; Northwest Caucasian)
Hewitt 1979

0) Basic categorial information

--- Noun: only a single "case", Adverbial/predicative -s (101; see [11] below); plural suffix (149); possessive-pronominal prefixes (102, 116)

--- Verb: codes Subj, DObj, IObj, and various obliques prefixally (101-3, 208ff.), normally explicitly doubling all full-NP arguments with a mark on the verb; ergative patterning, S/O vs. A (208-9):

Prefix-slot I (leftmost): S/O (intrans Subj, trans Obj)

Prefix-slot II: IObj (same markers for Obj-of-Postp, Possessor)

Prefix-slot III: A (trans Subj)

--- Article: definite-generic a-, indefinite -k' (152ff.), attached (respectively) to first and last element in NP (225-26)

1) Conjugated adpositions --- YES (103, 234)

--- Postpositions take series-II pronominal prefixes (which code IObj on verb, pronominal possessor on nouns); most postpositions require pronominal prefix copy even with full-NP object (234)

2) Word order

--- SOV (103), but flexible

--- Postpositions exclusively (103)

--- N-Adj, except for a few Adj categories where have Adj-N (222)

--- Gen-N (116)

--- RCl-N (35)

3) Relative clause linker --- ZERO

4) Relativization strategy/ies --- GAP-COPY

--- Coreferential noun is gapped; any role can be relativized so long as it is pronominally marked on Verb (or on Noun as a possessor); Obj-of-Postp not relativizable except by incorporating Postp into V (see [6] below) (36-38)

--- Verb undergoes 2 changes:

(i) Assumes a special "non-finite" form, used for subordination of all sorts (35, 201-2); however, "non-finite" is a misnomer, since verb-form codes same actants as finite verb (see [10] below)

(ii) Whichever verbal pronominal prefix is coreferential with Head N changes to one of two special "relative affixes", y(ə)- for series-I pronouns (S, O), z(ə)- for series II/III (A, IObj) (35); might be taken as analogous to "relative pronoun", since differentially encodes Noun's role within the RCl (though grammar book does not suggest this)

--- If relativized N is an embedded possessor, its coreferential prefix (on the embedded possessed N) changes to relative affix-II/III (37); this happens even if coref N is also marked on V (37), so coref N can be marked twice ("The boy WH did not go for a walk with WH mother [= with his mother]")

5) Special relative form of verb --- YES

- Presence of special "relative affixes" (35)
- 6) Polypersonal verb --- extremely (3+ actants)
 --- Codes S, DObj, IObj prefixally (101-3, 208-9); also can incorporate mini-postpositional phrases with pronominal object, [PronObj+Postp] (103-4, 209ff.)
- 7) Infixing/suffixing alternation --- NO
 --- Many directional preverbs; but their position is fixed between Slot-II and Slot-III prefixes (125)
- 8) Definite article in genitive embeddings
 --- Gen-HeadN; HeadN apparently must be marked with a pronominal copy of Possessor (116); this possessor affix occupies same slot-position as does definite-generic article a- (56), hence they're mutually exclusive
 --- Resultant pattern is: the-boys their-house (*the-their-house) (116); i.e., article only on embedded N
- 9) Nonconcord of V with full-NP Subj --- NO
 --- If the Slot-I prefix y(?)- (3-sg nonhuman, 3-pl) is immediately preceded by its referent (either as S or as O), prefix becomes zero (209, 211); but if the prefix is stressed, only y deletes, leaving schwa behind (268)
 --- This last fact suggests that we have here syntactically conditioned phonological change, not disappearance of a concord marker per se
- 10) Verbal abstract: VN or Inf? --- VN
 --- Basic nominalization ("masdar") changes any Slot-I prefix, either in S or O function, to a possessive (112); Slot-III prefixes (A) recast as instrumental (extraverbal); thus:
 your-seeing by-me = "my seeing you" (112)
 --- There is also a whole series of V-forms called "non-finite" (201-2), morphologically parallel to corresponding "finite" forms but with special suffixal coding, and used in all sorts of subordinate-clause contexts, including EQUI contexts (39ff. for "Adverb-clauses"; "Purposive" verb-forms (42-43, 199-201) are analyzed as a special "non-finite" subtype); however, as all these evidently encode tense and person/number more or less identically to normal finite forms, "nonfinite" is a misnomer: "nonfinite" will be taken here as finite
 --- For "Absolute" form used in Clause-chaining, see [15]
- 11) Predicative particle? --- YES (in certain constructions)
 --- Recall: the single "case" is Adverbial/Predicative -s ("I regard the woman as a fool") (101)
 --- Several copular roots, including zero; Subj NomPred (Cop)
 --- Three ways to make NomPred constructions (46-48, 106-8):
- (i) Zero copula; use NomPred (N or Adj) as a stative verb-root
 - (ii) Some copular verbs take NomPred in basic caseless form
 - (iii) One copula ("be, exist") takes NomPred in Adv/Pred -s
- Given the minimal role of case in Abkhaz (101), this ostensible Adverbial/Predicative "case" -s might well be recast as a "predicative particle"; not identical to any Postp

- 12) Postpositional periphrastic (VN + Postp) --- sort of
 --- One option for progressive is to create a stative verb using the Postpositional Phrase "in it" as verb-root, then using VN (masdar) of main verb as complement of this (181); thus:
 Art-play-masdar he-[it+in]-present = "he is playing"
 --- Not so much "He is [in playing]", as rather "He in-s playing"
- 13) "DO" periphrastic (VN + "DO") --- no mention, assume NO
- 14) Adverbial clause = "and" + finite clause --- barely
 --- Clause-level conjunction normally by clause-chaining, with -g'ə "and" (62, 242) possible as linker between chained clauses; only mention of "and" in adverb-clauses is to add nuance of "however much" to indefinite adverb clauses of manner (41), the latter involving "non-finite" (= finite, see [10]) verb form; here -g'ə "and" is better rendered "ever", as in "however" (perhaps compare German "wie Adj...auch" [? my suggestion])
- 15) VN/Inf instead of finite main-clause form --- apparently not
 --- Clause-chaining (218-19) typically uses either Past-Indef (174-75, 182, 201ff.) or Past-Absolutive (110-11):
 --- Past-Indef is a "non-finite" form (i.e. counts as finite, see [10])
 --- Past Absolutive is a normal finite form except that it omits Trans Subj-prefix (Slot III), retains Intrans (Slot I); despite Subj omission, this will be taken as fully finite, since S and O are coded in the normal way, as would be normal for an ergative language
 --- In any event, these forms appear specialized for use in clause chaining; neither fulfills any normal functions of "VN/Inf"
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- seemingly not
 --- No examples readily apparent in "Lexicon" section (273-81)

Afar (Northeast Africa; Cushitic [Afroasiatic])
 Bliese 1981

- 0) Basic categorial information
 --- Noun: case marked suffixally: nom, acc, gen (162ff.; 189 for pronoun paradigm); gender marked by stress and word-shape (180); plural marked suffixally (176-80); slot sequence (162); no article (15)
 --- NB: Unmarked form is Acc (zero ending); Nom only for Subj
 --- Verb: codes only Subj, either prefixally or suffixally (110)
- 1) Conjugated adpositions --- barely
 --- Most postpositions enclitic (3); grammar presents nominal and pronominal Obj-of-Postp on a par (46); several pronouns (3f.sg, 2.pl, 3.pl) occur in their long form when used as Obj-of-Postp (189-90)

--- But two pronouns (1.sg, 2.sg) assume short form before Postp if closed syllable created (e.g. *koo-t > ko-t "by you") (190); this shows a degree of phonological fusion

2) Word order

--- SOV (6)

--- Postpositions (45)

--- Adj-N (presumably); "Adj" is a special category of verb ("Modifier Verb"), with a distinct conjugation pattern (159); hence Adj-N should = RC1-N (note that the variant "N-RC1" will not occur here, as the RC1 contains no NP (see [4] below))

--- Gen-N (164); also N-Gen if postposition -k "of" is used and Gen is a pronoun (108, 170), e.g. "car us-k"

--- RC1-N (19); also N-RC1 (24, 106-8) with special suffix -iyva on verb

3) Relative clause linker --- ZERO (19)

--- Suffix -iyva not positioned appropriately to be linker

4) Relativization strategy/ies --- GAPPING; Postp moves

--- Gapping (exx. of Subj and Obj RC1, 21), order RC1-N; no further mark; "Modifier Verbs" (Adj) must appear in 3-sg (20)

--- Postpositional RC1s: N+Postp deletes, but just before the V there appears an adverbial element (elle "locative", edde "other") (21ff.) which (NB) can assume an anaphoric function in non-RC1 (23)

--- When HeadN has an additional (pre-Head) modifier in addition to the RC1, the sequence "Modifier-RC1-HeadN" is ungrammatical if RC1 contains a surface NP ("Modifier" would be mistaken as modifying the NP in the RC1, 107); rather, the different order "HeadN-RC1" is obligatory (24, 106-8); RC1 verb will then take special suffix -iyva followed by case marker, while HeadN will be followed by a conjunction; syntagm is thus:

Modifier HeadN-and [X V-iyva]-Case

--- Nothing appears to be said about genitival relativization; the genitive subtype in -k perhaps might come under Postp relativization

5) Special relative form of verb --- sometimes

--- Special suffix -iyva (for postposed RC1), termed a "relative pronoun" (24), perhaps could be taken as creating a distinct V form

--- NB: identical suffix (but with stress) forms verbal nouns from "Modifier Verbs" (i.e. Adjectives) (149)

6) Polypersonal verb --- 1 actant (Subj) coded (110)

7) Infixing/suffixing alternation --- not applicable

--- No Obj coded on V

8) Definite article in genitive embeddings --- not applicable

--- No article; but there is a "Particular" suffix -yta (175-76); semantically usually singular with definite connotations (176)

--- Gen syntagm: Possessor(gen) - Head (164ff.) ;

or can use Postpos -k "of, from":

Possessor-k Head (170)

or: Head Possessor-k (when Possessor is pronoun, 108, 170)

9) Nonconcord of V with full-NP Subj --- YES (though details puzzling)
 --- Fem-sg concord (110-11) when Subj is

- (i) Nominalized sentence
- (ii) Certain type of verb-noun
- (iii) Any conjoined Subj, even if both conjuncts are masc
- (iv) Most (?) plural subjects (122, 125)

--- Exact treatment of (nonconjoined) plural subjects confusing: f-sg concord with all nouns? (thus 125) or with nonhuman nouns? (thus 122); "Number in Afar verbs is only differentiated in connection with pronouns and human nouns" (122); yet exx. (126) show f-sg concord even with human plural Subj

--- Exact treatment of conjoined subjects confusing: conjoined subjects said to take f-sg concord (122, 125); but elsewhere (67) conjoined Subj NPs are presented as behaving in two ways (under particular conditions): either:

- (i) Both NPs take Acc (= unmarked case); if so, concord is f-sg (exx. 68, 96, 123, 174); OR
- (ii) Second NP is Nom (first is Acc); if so, full concord obtains (exx. 68, 96)

10) Verbal abstract: VN or Inf? --- INFINITIVE

--- Grammar presents two types of verbal abstracts, viz. "VN" (148-51) and INF (153-54), but no statement re verbal rection for either; see table (13) for nonfinite usages in embedded clauses (cf. also 31ff., 115-16, 153 for uses of "INF")

--- One ex. of "VN" with Acc Object (148); exx. passim of INF with Acc Object under discussion of INF as Verb-complement (36ff.; good exx. 32, 39, 44); hence both apparently pattern as "Infinitive"

--- Also have a category "NP Complement", whereby a normal finite clause takes a nominalizer suffix (16-19)

11) Predicative particle? --- not really

--- Normal construction: Subj Pred BE (7); but Copula often deletes (7ff., 94-95), and if so, a Cons-final PredNoun adds a final vowel; but this same vowel also appears in citation form of Noun, or before clitic Postposition; hence vowel is not really a "predicative particle" per se

12) Postpositional periphrastic (VN + Postp) --- YES

--- Special "engaged-in" construction (121-22), involving Postp -t meaning "in, at" (3); two variants, one being VN-t BE; ex. (122):
 We see-VN-t we-were "We were engaged in seeing"

13) "DO" periphrastic (VN + "DO") --- almost

--- Verbs can be changed into "Compound verbs" (146-47), whereby verb assumes a special Inf form which is followed by a helping-verb ("put" for transitive, "say" for intransitive); serves to emphasize the verb (147); construction exactly right except that helping verb isn't "DO"

14) Adverbial clause = "and" + finite clause --- no mention, assume NO

--- Sentence conjunction ("and, but") essentially a suffixed vowel

(62ff.), but no hint of any adverbial use

15) VN/Inf instead of finite main-clause form --- no mention, assume NO

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- no data

Albanian (Balkans; Indo-European)

Camaj 1984; also Newmark et al. 1982, Morgan 1972, Mann 1948

0) Basic categorial information

--- Noun: gender (masc/fem, 10ff.); number (sg/pl), plural form not always predictable (11); suffixal case (32); article (see below)

--- Verb: Subj coded suffixally (136-37); takes Obj clitic, frequently doubling a full-NP Obj via the characteristic Balkan "double object" construction (94); DObj and IObj clitics may cooccur, fusing together (95); many compound tenses with "be", "have", or Modal as Aux (126, 130)

--- Article: nouns have distinct Def and Indef declension, with case and definiteness essentially conflated in the Def declension (30-33); some nouns have double Def marking, with prenominal Def marker cooccurring with postnominal case/def suffix; prenominal marker shows a reduced case marking (38-39)

--- Two dialects (Geg, Tosk), with some grammatical differences

1) Conjugated adpositions --- NO

--- Explicit statement: Preps not inflected (63)

--- Pronouns appear in two series, full and clitic (93), the latter apparently only used in "double Obj construction" (94-95); no statement re pronoun type for Obj of Prep, but the single relevant ex. ("among us", 70) shows pronoun in non-clitic form

2) Word order

--- SVO (233, 239 exx.)

--- Prepositions (69ff.)

--- N-Adj usually (55); Adj-N for a few Adjs and superlatives (58)

--- N-Gen (41-43)

--- N-RC1 (1982:121, 1972:63)

3) Relative clause linker --- two types

--- Either invariant që, especially colloquial; or case-inflecting interrogative cili with preposed Article, functioning as RelPron (116)

4) Relativization strategy/ies --- GAPPING

--- Invariant që uses Gapping, like English that; but cannot strand Preps, hence që cannot be used with Prep RC1s (1972:64); one ex. of non-standard use with "genitival" RC1 (1982:279), with genitive recast as Dative:

ants që to-them is destroyed nest = "whose nest is destroyed"

--- Inflecting cili very like English WH-relatives, i.e. Gapping with

fronting of RelPron; exx. of Prep RC1 and Gen RC1 (1972:63); Prep and genitival Linker are required to front along with RelPron (63)

- 5) Special relative form of verb --- NO
- 6) Polypersonal verb --- 3 actants coded
 --- Suffixal Subj; clitic DObj, IObj (both together)
- 7) Infixing/suffixing alternation --- YES
 --- Obj clitics normally precede verb, in order: Ptc1-Clitic-Verb; but in Imperative, may optionally follow (94, 1982:26)
- 8) Definite article in genitive embeddings
 --- Syntagm (42-44): (Art)-HeadN-(Art) Linker (Art)-Dept-(Art); either/both NP(s) may take Definite declension if appropriate (per exx.)
 --- Linker looks just like forms of the Article; but apparently no categorial confusion, since Linker and Art can cooccur (for definite nouns)
 --- Dept is in genitive case (morphologically identical to dative); Gen must take Linker (1982:135); Linker agrees in case with HeadN (per paradigms, 42)
 --- Linker also occurs between (most) N and Adj in N-Adj construction (51ff.)
- 9) Nonconcord of V with full-NP Subj --- NO (235, 237)
- 10) Verbal abstract: VN or Inf? --- INF
 --- The only morphologically nonfinite forms are participles (1982:62); used (i) in compound verb tenses, and (ii) to form Gerund, Privative, Infinitive (Inf in Geg dialect only), when preceded by various particles (1984:165-67)
 --- These nonfinite forms (ii) may take direct objects (1982:93-94), apparently just like finite DObj (not in genitive, per exx.)
- 11) Predicative particle? --- NO
 --- Syntagm: Subj Cop Pred (with Pred in nominative, 236-37); Cop omissible (234); chapter on "particles" mentions no predicative particle (1982:310ff.)
- 12) Prepositional periphrastic (Prep + VN) --- very weakly
 --- "Privative" construction: pa + Participle = "without VERB-ing" (1982:64, 94-95; 1984:165)
 --- Privative may be dependent on some stative verb, inter alia "be" (1984:165, cf. ex. 1982:95)); but no indication that this really counts as a "construction"
 --- Note that Infinitive (Geg only) is formed by: me + Ptcpl (165); me is identical (or homophonous?) to the Prep "with"; but no indication that this construction occurs after verb "be"
 --- Gerund formed with nonprepositional particle tue (166)
- 13) "DO" periphrastic ("DO" + VN) --- apparently NO
 --- No mention in section devoted to "Verb formation" (205-13); Greek and Turkish loans are formed with special suffix (213)
 --- Construction běj ("do") + Subjunctive exists, indicating beginning

of an action (1982:106-7); but this is finite

- 14) Adverbial clause = "and" + finite clause --- weakly (semantics)
 --- In colloquial language, "and"-conjunctions may "join verbs in what appear to be subordinate relationships" (1982:302); 5 exx., most conveying purposive nuance, sometimes EQUI (cf. 1982:107)
- 15) VN/Inf instead of finite main-clause form --- not really
 --- Uses of nonfinite forms discussed (1982:64-65, 92ff.; 1984:165); only suggested parallel to main-clause use is use of gerund (tue + Ptcpl) to indicate a coordinated action (1982:98; ex. "<QUOTE>, he said, turning his back"); analogous to English, too weak to count as a real instance of intended construction
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- NO
 --- No good exx. in Mann 1948 (dictionary)

Amele (New Guinea; Gum family [Mabusu stock])
 Roberts 1987

- 0) Basic categorial information
 --- Noun: only inalienably possessed nouns show any inflection at all (149, 151): suffixed for pers/numb of noun's Possessor, sometimes (kin terms) also for sg/pl of noun itself (171); slot sequence is:
 N-possessor-pl ;
 nouns very often followed appositionally by coreferential pronoun signalling number (210)
 --- Verb: marked suffixally for Subj, for at most two Objs, and for +coref with next Subj (154, 276ff.)
 --- Article: only indefinite; postnominal particle (88, 203-4)
- 1) Conjugated adpositions --- NO (321)
 --- Free pronouns (table, 207-8) used as Obj of Postp
- 2) Word order
 --- SOV (70-71, 161), fairly rigid
 --- Postpositions (81ff., 156-57, 160-61)
 --- N-Adj (87, 97, 155)
 --- Gen-N (86, 213)
 --- Usually N-RC1, per reanalysis (externally headed, see [4])
- 3) Relative clause linker --- ZERO or invariant (218-19)
 --- Irrelevant on internal-head analysis; optional invariant marker eu on external-head reanalysis (see [4])
- 4) Relativization strategy/ies --- requires total reanalysis
 --- Grammar presents RC1s as internally-headed ("replacive", 49ff.),

but with the proviso that internal HeadN is obligatorily fronted within the RC1; RC1 optionally followed by demonstrative eu "that" (see 214-15), functioning as subordinating conjunction; the RC1 itself tends to occur initially within matrix clause; the (internal) HeadN may recur in matrix clause (in its normal position), but usually not; syntagm thus given as (49):

(i) [[HeadN ... Verb(rel)] (that) ... (repeated HeadN) ... V]

--- Structural indeterminacy apparent here: except where HeadN is repeated (which is not the norm, 49), this syntagm seems to me equally analyzable (and much more simply) as a straightforward externally-headed RC1 (recall there is no case marking):

(ii) [HeadN [... Verb(rel)] (that) ... V]

--- Only where HeadN is repeated does this fail to work; but here, an internally headed analysis is incompatible with repetition of HeadN; the attested structure would seem to demand analysis either as correlative (with fronted head, perhaps like Anglo-Saxon), or perhaps as externally headed [N-RC1] syntagm, fronted as topic and then repeated in main clause

--- See also remarks on temporal RC1 (below), which must be taken as externally headed

--- So the grammar's analysis works nowhere

--- All NP-types are asserted to be relativizable (52); but exx. often do not work: the sentences presented as exemplifying Accompanier RC1, Instrument RC1, and (one ex. of) Possessor RC1 really are Subj RC1 or Obj RC1

--- Possessor RC1 is allowed; a proviso is stated that alienable possession relativizable only as nonsubject of RC1, inalienable possession only as subject of RC1, but Roberts's putative example of the latter isn't right; only one real ex. of possession (50, 52):

(iii) man he of chicken boy he.stole (man) (that) he.comes
"The man whose chicken the boy stole is coming"

--- If HeadN functions as temporal, locative, or instrument in matrix clause (49), RC1 is followed by Postp na "at, in, by, with" (added after eu "that"); relativization is asserted to be possible (52) on temporal, locative, and instrument in the RC1, but his temporal and instrumental exx. (and a locative ex. on p. 57) involve these roles in the matrix clause; the locative ex. on p. 52 seems bona fide, but structurally puzzling, with Postp apparently moved to postverbal position from its inherent position after "house" (analysis?):

(iv) [[house ____ she food cooks] that in ... fell]
"The house that she cooks food in has fallen down"

--- NB: for temporal RC1, the HeadN sain "time" usually occurs directly after (not before) the RC1 (56; counterexample on p. 52):

(v) [... Verb(rel)] sain eu na... "at the time when..." ;
to analyze this as internally headed or correlative would require postverbal sain to be taken as falling within the embedded clause, yet Amele clauses of all types are standardly verb-final (71) (except for afterthought, 148); this type would thus seem to demand an external-head analysis [my own argument]

--- Tentative reanalysis: Take basic Amele RC1 as externally headed,

usually N-RC1, sometimes RC1-N; alternate correlative strategy if HeadN is repeated; take the one embedded-clause oblique RC1 ex. (iv) as representing Adp movement-to-V; take the one possessive ex. as involving resumptive pronoun (Copying):

(iii) man [[he-of chicken] boy he.stole] (man) (that) he.comes ;
but Postp RC1 and Possessor RC1 rest on very thin evidence

5) Special relative form of verb --- NONE

6) Polypersonal verb --- 3 actants maximum (suffixally)

--- Subj marker normally obligatory (276)

--- Some verbs must code an Obj, even if intransitive (281, 291); others cannot code an Obj, even if transitive (281, 286); others code Obj optionally (281, 284)

7) Infixing/suffixing alternation --- apparently NO

--- All verbal morphology suffixal (153); slot sequence seems fixed (exx. 279-81)

8) Definite article in genitive embeddings --- not applicable

--- No definite article; no data readily apparent regarding Gen-N with indefinite article

--- Inalienable possession: Possessor (CorefPron) HeadN (86), with HeadN coding pers/numb of possessor and perhaps its own number

--- Alienable possession: [Possessor na] HeadN, with Postp na "of" (213, ex. 47)

9) Nonconcord of V with full-NP Subj --- NO

--- Nouns seldom code plural anyway; any agreement between V and N thus usually via free-pronominal copy of the N (161-62, see [0] above); syntagm:

N CorefPron Verb (Ex.: man they they.came "Men came") (162)

--- Verb obligatorily codes Subj pers/numb (291); no hint of nonconcord; found no exx. where plural-coded noun was Subj

--- With conjoined Subj (using ca "with"), verb agrees with conjoined Subj as a whole (105, 169); with counted nouns, verb likewise shows agreement (157)

10) Verbal abstract: VN or Inf? --- INFINITIVE

--- Various "nonfinite" forms lack tense (272ff.); but only the form called "Infinitive" lacks Subj pers/numb (277), though it may code Obj (291, 292)

--- Inf can be marked with Obj suffix, just like a finite verb (292); exx. with full-NP Obj show normal (i.e. non-genitive) rection (47, 58)

--- Inf ending is -ec/-oc (272), identical to ending used for nominalization (47, 325) and adjectivalization (87, 327)

11) Predicative particle? --- NO

--- No "be" copula (64); can use verbs "lie", "stand", "sit", etc. as copulas, or verbless equational sentence (66); nominal or adjectival complement not marked in any special way (64)

12) Postpositional periphrastic (VN + Postp) --- no mention, assume NO

--- Durative aspect expressible as a serial verb-string ending in

"sit" or "lie" as copula (256), with only final (copular) verb marked for pers/numb; but no Postp, nor is copular verb a special "Locative" copula

--- Inf can occur as Obj of Postp nu "for" and we "like" (166-67); but not the right construction

13) "DO" periphrastic (VN + "DO") --- almost

--- There is a particular serial-verb combination: VERB + "hit" (311); usually has idiomatic meanings, but also used to code Obj (288) for a verb of the type which may not itself take Obj markers (see [6] above); in the latter use, the periphrasis seems synonymous with simple verb, hence exactly the right construction except that helping verb isn't "DO"

14) Adverbial clause = "and" + finite clause --- not applicable

--- Clause-level "and" conjunction ca (= the Postp "with"), usable only with Inf or nominalized Verb (98); hence no finite "and" conjunction; rather, usually use coordinate medial verbs (100ff.)

--- Adv clauses use various devices, listed explicitly (56), but not "and" conjunction

15) VN/Inf instead of finite main-clause form --- barely

--- Stand-alone Inf can be used with mood particles to express Hortatory and Certain-Apprehension moods (267-69, 272); no other independent uses of Inf, apparently

--- All clause-chaining forms (medial verbs) do have Subj explicitly coded (293)

--- However, when several verbs are strung together to make a serial verb (necessarily with same subject), non-final verbs occur in a "stripped-down" version of Sequential Same-Subj form (314); this latter has two variants, one of which deletes Subj marker (273); but serial verbs ought to count as a single verb nucleus, hence this doesn't look like the relevant construction

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- seemingly not

--- Only one barely relevant ex. in "Lexicon" section (380-90):

boy/child navel = "umbilicus" (384)

and one elsewhere: road its-mother = "main road" (330)

Amharic (Ethiopia; Semitic [Afroasiatic])

Leslau 1967; also Cohen 1936, Hartmann 1980; Leslau 1976

0) Basic categorial information

--- Noun: covert masc/fem gender (117); suffixal plural (41), suffixal DefArt (47-48), accusative suffix -n for definite objects (66-67), possessive pronoun suffixes (151ff.); slot sequence:

N - pl - {Art, poss} - n (48, 67, 154);

but if Adj precedes noun, Art and -n appear on Adj, not Noun

--- Verb: Subj coded prefixally (Perfect, 65) or suffixally (Imperfect, 107); Obj coded suffixally (209ff.), often cooccurs with a full-NP object (212-13), sometimes "infixes" (see [7]); Obj suffix may be preceded by an incorporated Prep, only possible with the Preps lä "to", bä "with, against" (276-79); thus:

VERB - [(Prep)-Obj]

--- Article: DefArt suffixal (47-48), but not obligatory for expression of definiteness (1936:108); Indef Art is a separate word, pronominal and optional (1967:37)

1) Conjugated adpositions

--- The standard series of pronominal suffixes can only serve as noun possessor (151) and verb Obj (212) and Subj of gerundive (287-88), though forms aren't precisely identical; series does not occur on free Preps, only on Preps lä and bä when incorporated into the verb (276-77)

--- Rather, Preps are said to take Indep pronouns (1936:136); but numerous exx. (1936:136, 1980:327ff. passim) show that initial vowel of Indep pronouns (paradigm 1967:35) elides when combined with Prep, e.g.

wädä + əne > wädäne "to me" (ex. 1980:335) ;

Leslau explicitly notes (regarding Prep yä- "of") that even forms spelled e.g. yä-?ənnat are pronounced yännat (1967:137); forms thus involve fusion, should therefore count as conjugated Prep

2) Word order

--- SOV (68), always verb-final

--- Prepositions; many postpositional elements, most cooccurring with Prep to yield a Circumposition, some occurring alone (67, 77-78; 1980:326-40)

--- Adj-N (38)

--- Gen-N (136)

--- RCl-N (332)

3) Relative clause linker --- ZERO

--- RCl marker is yä- (for verbs in Perfect), yämm- (for Imperfect) (332-33); but occurs prefixed to clause-final verb, thus not a real "linker" (not medial between RCl and HeadN); yä- is multi-purpose subordinator, also functioning in genitives (see [8])

4) Relativization strategy/ies --- COPYING

--- Syntagm: [RCl ... yä(mm)-VERB]-Art-n HeadN (335-36); relative marker yä- for Perfect, yämm- for Imperfect; syntagm may occur headless (1936:114)

--- If HeadN functions in matrix clause as Obj-of-Prep (or as Obj of genitive Prep yä- "of"), then that Prep will appear inside RCl, overlaying and replacing yä- (336, 1936:115) (compare [8] below); if verb is in Impf, then Prep overlays just the yä-, changing yämm- to Prep-mm- (1936:115)

--- Subject RCl gaps; all others leave pronominal copy (1967:336; especially clearly 1936:117-18)

--- As for prepositional RCl, if Prep is lä- or bä-, PrepPhr [Prep+PronObj] is incorporated into verb (with pronoun Obj); for several other Preps, PrepPhr deletes but "reincarnates" within verb as incorporated PrepPhr [lä- or bä- + PronObj]; still other Preps survive in situ with resumptive pronoun (1980:340-48)

- 5) Special relative form of verb --- weakly
 --- yä+Verb also used to mark verb in clausal object of "think" (346); marker yä- used also for genitives (see [8] below) (136)
 --- Marker yämm- as such is unique to RCl; but without the y, it's used with several conjunctions (346); syntagm: Cju-yämm-Impf
- 6) Polypersonal verb --- 2 actants coded
 --- Subj prefixally or suffixally, Obj (optionally with Prep lä-, bä-) suffixally
- 7) Infixing/suffixing alternation --- weakly
 --- "Compound Imperfect" tense formed from Impf + allä "exist" (225); here Obj suffix will come between main verb and allä (225ff.); ex.
 yänägr - äw - al "he tells him" ;
 Hartmann calls this an "Objektinterfix" (1980:174ff.); but could reanalyze with a final slot (post-Obj) optional
 --- NB: This tense-form spelled and treated as a single fused word
- 8) Definite article in genitive embeddings
 --- Syntagm: yä-Dept HeadN (136); if NP as a whole is governed by a Prep, yä- is omitted and overlaid by that Prep (137) (compare [4] above); thus too in a double genitive (1936:78, 1980:395):
 *yä-[yä-Noun1 Noun2] Noun3 > yä-Noun1 Noun2 Noun3
 --- If NP as a whole is Def, article goes on Dept (1967:136):
 yä-boy-Art book = "the boy's book" ;
 general rule (1980:306) is that Art goes on pronominal modifier in NP, whether Adj or RCl or Gen (recall [0], [4]); but if Dept+HeadN is taken as a compound, Art is suffixed to entire compound (1936:87)
 --- NB: one example (mentioned without comment in passing) seems to show two articles (1980:390):
 yä-man-Art palm-Art = "the man's palm"
- 9) Nonconcord of V with full-NP Subj --- apparently NO
 --- Full Subj-Verb concord is the rule (1936:159, cf. 1967:55, 64); minor optional nonconcord noted with counted nouns (1936:287), or "avec plusieurs sujets exprimant des abstractions" (1936:159; no further explanation)
- 10) Verbal abstract: VN or Inf? --- Inf
 --- Verbal abstract in mä-, described as equivalent of English infinitive or gerund (1967:253, 1936:194); full-NP Obj precedes verbal abstract, can take -n of definite Obj (ex. 1980:382), does not take genitive yä-; i.e., rection same as for finite verb, hence Infinitive
 --- Pronominal Obj of Inf seems not to occur as Obj suffix (no explicit statement), but rather as Indep Pron (1936:135, 194); Subject of Inf can be expressed as possessive pronoun suffix (1967:253)
 --- Gerundive: a nominal form used in clause-chaining, fundamentally an adverbial (1967:288, 1980:195, 1936:183); Subj marking obligatory (via possessive suffix), followed by normal Obj pronouns (1967:290); rection of full-NP Obj same as for finite verb (ex. with definite -n, 1967:289)
- 11) Predicative particle? --- NO
 --- Copular syntagm: Subj Pred Cop (37); distinct Cop and existential

verb; Cop omissible in proverbs etc. (1936:355)

12) Circumpositional periphrastic (Circum + VN) --- YES

--- Circumposition bä ... lay "on"; used with Inf to express Progressive (1980:207, 208, 370), in syntagm:

bä-Inf lay COP e.g. "Er ist beim Lesen"

13) "DO" periphrastic (VN + "DO") --- almost

--- Amharic has composite verbs (1967:363, 1980:141-43; best discussion 1936:262ff.), with noninflecting element X preceding conjugated Aux; Aux is usually "say", sometimes "do" ("do" makes verb Causative, 1980:141)

--- Element X may be onomatopoeic, or some other element that occurs only in composite verb; but additionally, any verb-root has a special form (usually with final geminate) specifically for use as "X" in composite verbs; composite verb has expressive nuance, adds intensive or reductive semantics

14) Adverbial clause = "and" + finite clause --- weakly (semantics)

--- Conjunction -ənna "and" operates both at clause-level and NP-level; per Cohen, not semantically restricted to "une simple liaison", but also "donc, car, en effet, puisque, comme", almost like "une subordination lâche" (1936:310; cf. 1980:416-17)

--- Adv-clause-equivalent more typically formed by Gerundive (1967:289) or various adverbial conjunctions (1967:399-475 passim) or RCl on time/place HeadN; cf. 1980:427ff.

15) VN/Inf instead of finite main-clause form --- NO

--- Gerundive used heavily in clause-chaining (1967:289, 1936:183, 1980:195), but always encodes Subj (as possessive suffix) (1967:287), hence not "nonfinite" as used here; Inf apparently has no main-clause use except in proverbs etc. (1936:355)

16) Word-initial change --- NO

17) Extended use of kin terms ("Kin of Noun")

--- Only exx. in Leslau 1976 are: "mother of key" = lock, "father of custody" = guardian, "father [or child] of breast" = protector [or protegee] (formal relationship)

--- Also əmmā-bet "mother of house" = mistress of the house; but here "mother" is the Geez (not Amharic) word

Babungo (West Africa [Cameroon]; Grassfields Bantu [Niger-Congo])
Schaub 1985

0) Basic categorial information

--- Noun: affix for noun-class/number (133, 171ff.), usually prefixal but one suffixal (172); remnants of a locative case (marked tonally) (65, 69)

- Verb: no person/number marking (232)
- Article: grammar says no definite article (73, 192); but have an "anaphoric demonstrative Adj" ghō (agrees with noun-class of HeadN), described as corresponding to anaphoric use of Engl. definite article, and obligatorily following any noun already mentioned in text (97, cf. 203); will reanalyze as "hedged article"
- 1) Conjugated adpositions --- NO
 - Prepositions, noninflecting (134, 237)
- 2) Word order
 - SVO (62)
 - Prepositions (237)
 - N-Adj (77)
 - N-Gen (77; exx. 75, 139, 146)
 - N-RC1 (33, 77)
- 3) Relative clause linker --- Rel Particle
 - Two invariant relative particles, in free variation (32), positioned between Head and RC1
- 4) Relativization strategy/ies (32ff.) --- usually COPYING
 - Can relativize on any role in RC1 (36); relativized constituent remains in place, replaced by pronominal copy (34); only DObj usually undergoes deletion (34); note that Subj does get copied
 - For restrictive RC1, Head-N is modified by a demonstrative ("that boy who..."); sequence is thus:
 - HeadN (Dem) RelPtcl [RC1]
- 5) Special relative form of verb --- NO
- 6) Polypersonal verb --- zero actants coded (232)
- 7) Infixing/suffixing alternation --- not applicable
- 8) Definite article in genitive embeddings --- hedged
 - Construction: Head and Possessor separated by "Associative Marker":
 - Head - Ass.Mk - Poss (146, cf. 75);
 - Ass.Mk agrees with Head in noun class, but often zero (with tonal effects) (139)
 - Take the "anaphoric demonstrative Adj" as "hedged article"; but no exx. of Possessor are presented involving this anaphor
- 9) Nonconcord of V with full-NP Subj --- not applicable
- 10) Verbal abstract: VN or Inf? --- INFINITIVE
 - Two nonfinite verb forms (defined by absence of tense/asp/mood) (231-32):
 - (i) Form used in purpose clauses (= "in order to"; prefix mā-), used after Prep māa "for", exx. show Obj in normal form (unpossessed); also used as a nominalized form (single ex.: "I have heard his thanking")
 - (ii) Zero-marked form, used only in "closely knit verb strings"

(serial verb constructions, apparently); exx. show Obj in normal form
 --- Explicit statement that NPs are identically marked with nonfinite
 and with finite verbs (144)

11) Predicative particle? --- NO

--- Copula is obligatory with PredNoun and PredAdj (48); no particle;
 construction is:

Subj Cop PredNom (49-50)

12) Prepositional periphrastic (Prep + VN) --- NO

--- Only use of verbal abstract with Prep is: máa ("for") + "Purposive" (see [10] above); not a main-clause tense construction

13) "DO" periphrastic ("DO" + VN) --- no mention, assume NO

14) Adverbial clause = "and" + finite clause --- not applicable

--- No conjunction "and" (78); Adv Clauses made with subordinating particles (37ff.)

15) VN/Inf instead of finite main-clause form --- not really

--- The one possible candidate is the zero-marked type of nonfinite form (see [10] above), which appears only in "closely knit verb strings" (on non-initial verbs) (232); but such verbs would appear to constitute a single complex verb (only the first in the string takes Perf/Impf aspect marker, 232), hence the nonfinite form isn't really standing in for an entire clause

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- some

--- Common use of wée "child" in what is described as a diminutive function (thus: small stone = "child stone", non-figurative use); but see these figurative exx. (248-49, 391-92):

finger	=	"child hand"	toe	=	"child foot"
Adam's apple	=	"child throat"	clitoris	=	"child vagina"

also: leopard = "father night"

very big thing = "mother thing"

(no other examples of "mother" as augmentative)

--- These are presented as compounds, not as possessive NP constructions, but often impossible to distinguish the two structurally (248); when order is "Modifier-Head" it contravenes normal syntactic ordering and hence must be a compound (248), and this is claimed to hold for the "child" exx. above (thus in "child stone"); but might analyze (e.g.) "child hand" as "child of the hand" (i.e., "that of the hand") rather than "small hand"; hence unclear what counts here as conceptual head, and whether diminutive analysis is the best one

Basque (Spain/France (Pyrenees); isolate)
 Saltarelli 1988; also De Rijk 1972, Lafitte 1962, Rebuschi 1984
 [My thanks to Begotxu Olaizola for advice and insight on many points]

0) Basic categorial information

--- Noun: inflects (suffixally) for case (rich system), number (146), and definiteness (200-201), all more or less conflated (200); no gender; Ergative-Absolutive case marking

--- Verb: obligatory pers/num marking for Subj, DObj, IObj (238ff.); most verbs inflect periphrastically, with an inflecting Aux preceded by participial form of main Verb (Intro iii-iv); Aux codes pers/num, tense, mood; ptcpl codes only aspect

--- Article: no separate word; nouns inflect as def-sg, def-pl, or indef (200ff.; cf. paradigms, 300); semantics not exactly "definiteness" (Rebuschi 123ff., 128-30)

1) Conjugated adpositions --- NO

--- All pronouns are free (205ff.), take full case inflection (paradigm: Lafitte 1962:89); many of the case endings are called "postpositions" (Saltarelli 72), here reanalyzed as "hedged postpositions"; additionally, there are independent postpositions (251)

2) Word order

--- SOV (66-67), but much freedom; OV more rigid in subordinate clause

--- Postpositions (19, 72, 144, 251)

--- N-Adj (80)

--- Gen-N (80)

--- RCl-N or N-RCl (36ff., 80)

3) Relative clause linker --- ZERO (also Relative Pronoun)

--- "Complementizer" -(e)n, attached to RCl-final V (36); analyzed as part of verb form (see [5])

4) Relativization strategy/ies --- GAPPING

--- Basic strategy: add complementizer -(e)n to verb-final finite clause; this either precedes Head (with no case marking on RCl) or follows Head (with both Head and RCl-verb agreeing in case/numb/def) (36); coref N is gapped; strategy may be used if coref N in RCl is Erg, Abs, or Dat, or if coref N is Oblique and fills identical roles in matrix clause and RCl (37)

--- Second strategy: same as first, but add marker -ko to RCl formed by first strategy (thus -(e)n-ko); this enables relativization on other obliques (38); same ko is used to form all sorts of adnominal complements (161) (e.g. "the boy [with a good heart]")

--- Third (purely literary) strategy: RCl-initial WH-pronoun (= interrogative), which is case-marked to show role in RCl; RCl follows an independent clause containing coref N, and is marked by -(e)n (or another subordinator); description unclear, but appears to be correlative strategy (38-39); will be ignored here

--- No explicit discussion of (nonliterary) Possessor RCl; can be formed only by recasting the RCl as non-genitival, roughly: "the boy [to-whom the father died]" [BO, p.c.]

--- No discussion of (nonliterary) postpositional RCl's involving

independent Postps

- 5) Special relative form of verb --- not exactly
 --- Marker -(e)n used for all sorts of subordination (30ff., 43ff.)
- 6) Polypersonal verb --- 3 actants coded suffixally (Subj, DObj, IObj)
 --- Ergative system; affix ordering basically: Abs-Stem-Dat-Erg; some split ergative marking when 1/2-pers Subj acts on 3-pers Obj
- 7) Infixing/suffixing alternation --- NO (apparently no preverbs)
- 8) Definite article in genitive embeddings
 --- Syntagm is: [Possessor-gen] Head (also if Possessor is a Pron) (76, 161)
 --- No explicit discussion, but some exx. (161, also Rebuschi 114ff.) show Def marking on both NPs, which apparently is standard [BO, p.c.]; for morphophonology see Rebuschi 151
- 9) Nonconcord of V with full-NP Subj --- NO
 --- All verb agreement obligatory, uninfluenced by word order (244); verb agreement-marker can be sg or pl when NP is quantified-indefinite (245)
- 10) Verbal abstract: VN or Inf? --- usually INFINITIVE
 --- Basic productive nonfinite nominalizer ("corresponding to gerunds", 154) is -te (30, 154, 258), governs same cases as finite verb (154; cf. 30, 258); however, in northern dialects, DObj (NB: not intrans Subj) can be marked with genitive instead of Abs (155), which would count as VN
 --- Other nominalization suffixes take Subject in genitive; DObj also in genitive, or forms a compound with verb; these are lexically governed (156, 258-59), appear to be more like derived nominals
 --- Also have "absolute" (participial) construction (153-54), with adverbial or nominal function; DObj as with finite verb
- 11) Predicative particle? --- NO
 --- Copula always present in independent copular clauses (61); PredNom in Nominative, agrees with Subj in number, and normally definite (150; cf. Rebuschi 128); no particle mentioned; order is:
 Subj PredNom Cop
- 12) Postpositional periphrastic (VN + Postp) --- sort of
 --- Progressive formed from locative case of the te-nominalized main verb, followed by verb meaning "be engaged in, continue"; in some varieties, locative nominalized verb is instead followed by egon "be" (229):
 [music listen-te-loc] he.is "He is listening to music"
- 13) "DO" periphrastic (VN + "DO") --- YES
 --- Numerous combinations of "Obj+V" that act as "verbal units" (246); very commonly this main verb is "DO" (including one onomatopoeic ex., 290); however, almost always the Obj is a primitive noun, not a deverbal nominal [BO, p.c.]; a very few exx. (294ff.) do appear to involve a deverbal noun ("bite, vomit"; also "play" [BO, p.c.]

--- Also: In some dialects, can insert "DO" between participle and conjugated Aux to indicate "that it is the action... specified by the verb that is the focus of the sentence" (247)

--- A similar "DO" insertion commonly occurs in short confirmatory responses to yes/no questions; the questioned verb (but not the remainder of the question) is repeated, with "DO" added [BO, p.c.; apparently not in Saltarelli]

14) Adverbial clause = "and" + finite clause --- weakly (semantics)

--- Conjunction eta "and" (82) can also be used as enclitic "because" (46-47), suffixed to finite clause; no pause before enclitic, slight pause before conjunction (83)

15) VN/Inf instead of finite main-clause form --- weakly

--- In verb chain of periphrastic verbs, Aux only appears after the last of the participles (246-47), hence most verbs appear in a nonfinite form; but this is not a "VN/Inf"

16) Word-initial change --- not really

--- Diminutive can be coded by changing a consonant (dental or guttural) to palatal equivalent (269-70, 276-77, Lafitte 147); highly lexicalized, not the only diminutive technique (270, Lafitte 147-49); the consonant affected need not be word-initial or even the first consonant in the word (per exx., 270, Lafitte 147)

17) Extended use of kin terms ("Kin of Noun") --- NO [BO, p.c.]

--- No examples readily apparent in "Lexicon" section (291-99)

Burushaski (northernmost India (Kashmir); isolate)
Lorimer 1935

0) Basic categorial information

--- Noun: 4 genders (14ff.), usually covert (17); various plural suffixes (25ff.); singulative suffix -an [my term], often used as indefinite article (46-50); case suffixes (list: 55, forms: 53ff., functions: 63ff.), split ergative pattern (Erg-Nom in Past, Nom-Nom in Present (no distinct Acc form)) (63-64), multiple case suffixes may cooccur (82); pronominal prefixes (forms: 127) indicating possessor with inalienable possession (133-37), also have other uses (130-31), e.g. with Verb and Postps; order of suffixes is:

Noun - an - Case (61)

--- Verb: suffixal inflection for Subj (paradigm 244-46); some verbs take pronominal prefixes (= Noun possessive prefixes), coding Obj (usually) of trans, Subj of intrans (192, 218-20, 264ff.), hence possibility of double coding of Subj (Pfx and Sfx); for many verbs starting with d-, these prefixes are instead infixes (192, 274-75), see [7] below; some verbs show change in initial consonant, displaying agreement with gender of Subj (intrans) or Obj (trans) (202-4); several compound verb tenses, built on Present or Past stem + "BE" (193, 244ff.)

--- Article: distal demonstrative (which may be either adjectival or pronominal, 140) frequently corresponds to English Def Art (146) [OG: notably in relative clauses], will analyze as "hedged article"; singular suffix can indicate indefiniteness, but optional (47)

1) Conjugated adpositions --- YES

--- Difficulty of defining class "postposition" (53-54); of 17-odd Postps (lists 62-63, 95-104, 131-32), most take pronominal prefixes as Obj (131-32 and exx. passim), only 3 (?) take Indep pronoun

2) Word order

--- SOV (402), fairly strict
 --- Postpositions (53, 62)
 --- Adj-N (119, 403)
 --- Gen-N (404, exx. 68-69)
 --- RCl-N (406)

3) Relative clause linker --- ZERO

--- No relative pronouns (176)

4) Relativization strategy/ies --- GAPPING, COPYING (Gen)

--- Exx. of RCls (all unglossed): pp. 118, 177, 333-40, 395-99, 406
 --- Normal strategy: participial, using "Static Participle" (forms: 299-303, functions 332-49); StPpl appears to have minimal person inflection (1-pers vs. other) (299), but usually occurs in 3-person (300)
 --- Usual syntagm is: (Dem) [RCl] (Dem) HeadN (335, 395-98), with RCl-final verb in StPpl; Demonstrative usually follows RCl, may occur both before and after (335, 404), and in one ex. (336) only before:

Dem [his-son bury-StPpl] field "the field where his son was buried"
 (Dem agrees here in gender with "field")

--- StPpl construable actively or passively (333-34), with coref N (Subj or Obj) gapped (no explicit statement, but many exx.); also HeadN can be noun of time or place (exx. 333, 336, 340, 396); no clear exx. (?) of Postpositional RCl (as opposed to time or place RCl)

--- Two exx. of genitival RCl, showing Copying (no explicit statement):

[his-father NEG.BE-StPpl] orphan "fatherless orphan" (334, 348)
 [his-beard white become-StPpl] man "man whose beard has become white"
 (340)

--- Another option: may use Inf instead of StPpl (360, 398)

--- Another strategy: correlative RCls ("indefinite relatives"):

[... [which/whoever HeadN] ... Finite.Verb] kE [Main.Clause] ,
 where main clause includes a resumptive pronoun (177-78, 398-99)

5) Special relative form of verb --- NO

--- Static Participle has other uses besides RCl (333-34): Adverbial; as PredAdj with "BE"; as a kind of gerund (342-46); as optative (341)

6) Polypersonal verb --- 2 actants coded

--- Subj suffixal (244ff.); Obj of trans, Subj of intrans verb coded

via pronominal prefixes (192, 218-20), but only certain verbs apparently may take prefix (218)

7) Infixing/suffixing alternation --- YES (but infixing/prefixing)

--- Verbs starting with initial separable d- (meaningless, 226) take Pron prefixes as infixes, between d- and remainder of verb root (224, 274ff.)

--- Also, the non-d- verbs "eat", "throw down" take infixed Obj markers (203, 222-23)

8) Definite article in genitive embeddings --- hedged

--- Syntagm: a) Dept-gen Head (alienable)
 (per exx. 68-69) b) Dept-gen PronPfx-Head (inalienable)
--- Taking Demonstr as "article", note these exx. (69, 143):

[that woman]-gen her-daughter-Indef "a daughter of the woman"
[old woman-Indef]-gen house-gen door "door of an old woman's house"
[this my-father]-gen these bones "these bones of this my father"

with the last ex. showing Dem occurring on both Nouns

9) Nonconcord of V with full-NP Subj --- no mention, assume NO

10) Verbal abstract: VN or Inf? --- INF

--- Grammar presents a form called "Infinitive" (form: 193-94, functions: 350-63), used in all sorts of prototypical infinitive functions

---- No explicit statement re Object rection, but all exx. (350-59) have Obj either as (a) Noun in Nom/Acc (not Gen), or as (b) PronPfx on Noun, the latter surely to be taken as Nom/Acc (not Gen) because PronPfx as Gen apparently restricted to inalienable possession

--- A unique ex. cited of Genitive rection of Obj, with Nom/Acc rection also possible (Obj in both appears as Indep pronoun):

my kill-Inf (69, 351) vs. me kill-Inf-Dat (69, 353)

--- Static participle (see [4] above) similarly shows Nom/Acc rection

11) Predicative particle? --- NO (or barely)

--- Basic syntagm: Subj PredN/PredAdj BE (402)

--- PredAdj may often take the singulative -an (48, 110, 119), as if to recast "it is black" as "it is a black one"; but this should hardly count as a "predicative particle"

12) Postpositional periphrastic (VN + Postp) --- no mention, assume NO

--- Many compound tenses with "BE" (193, 244ff.); also Inf + "BE", in sense of obligation (357ff.); Static Participle + "BE", in stative/adjectival sense (335-36, 338, 340); a few exx. of Past Participle Active + "BE", as Imperfect (331); but none of these involves a Postposition

13) "DO" periphrastic (VN + "DO") --- in a sense

--- Have many "compound verbs" (226-38) of form: Adj/Adv/N + MainV, especially if initial element is foreign (227), notably Persian or Arabic; MainV usually "do" or "become", but others also possible (234ff.)

--- But no sign that initial element itself can be a Verb form

- 14) Adverbial clause = "and" + finite clause --- perhaps YES
 --- "And" conjunction is kE, but rarely occurs as clause connector (381), since first verb usually expressed participially; language "is averse to a series of finite verbs" (330), but no statement of total exclusion; also have a distinct "and-then" conjunction (375-76, 382)
 --- "When"-clause can be expressed as [Finite-clause]-kE (385)
 --- Adverbial clause-equivalents discussed (384-95); Present Participle used for contemporaneous continuous action (328-29)
- 15) VN/Inf instead of finite main-clause form --- weakly
 --- All but last of a chain of parallel clauses are normally expressed participially (330, 381); Past Participle Active (384) is the usual chaining form, especially for same-subject
 --- But Past Participle Active apparently fulfills no standard Inf/VN functions (discussion of uses, 330-32)
- 16) Word-initial change --- no mention, assume NO
 --- "Initial change" (see [0]) merely indicates gender agreement
- 17) Extended use of kin terms ("Kin of Noun") --- NO
 --- No examples apparent in Lorimer 1938 (large dictionary)

Chinese (Mandarin) (East Asia; Sino-Tibetan)

Li & Thompson 1981

[My thanks to Wen Hsu for answers to several queries]

- 0) Basic categorial information: Isolating
 --- Noun: no cases; only "inflection" is plural -men, optional and only for human nouns (12, 40); no articles
 --- Verb: no person/number (12); no tense, yes aspect (suffixal) (13)
- 1) Conjugated adpositions --- NO
 --- Prepositions (called "coverbs", 356ff.); also Postpositions (called "locative particles", 25, 390ff.), usually cooccurring with Prep zài "at" to form Circumposition: zài NP Loc.Ptcl
 --- Term "coverb" because Chinese Preps typically are derived from verbs and often can be used as verbs (360)
- 2) Word order
 --- SVO (23: pragmatically basic order, textually more common; near-universal with complex sentences, 24); SOV also common (160ff., 463ff.)
 --- Prepositions (coverbs) and Postpositions (locative particles)
 --- Adj-N (117, 124); most Adjs are "adjectival verbs" (142)
 --- Gen-N (113); usually Gen-de Head, sometimes no de (115)
 --- RCl-N (116, 124)
- 3) Relative clause linker --- invariant particle
 --- Nominalizer -de, in syntagm: [Clause] -de Head (116-17, 579ff.)

- 4) Relativization strategy/ies --- GAPPING (sometimes Copying)
 - Gapping when coreferential N in RCl is Subj, DObj, various obliques (582-83); if the gapped coref N is oblique, not always straightforward what the equivalent non-relative clause would be: coref Noun would sometimes be accompanied by a (deleted) Prep, sometimes by a (deleted) serial verb [WH, p.c.; grammar makes no explicit statement]
 - Copying when coref N is IObj, or after certain Preps (584-85); characterized as correct but rare, somewhat awkward, hence "marginal" (585)
 - Possessor relativization: copying possible, but often very marginal [WH, p.c.; grammar does not mention at all]
- 5) Special relative form of verb --- NO
- 6) Polypersonal verb --- zero actants
- 7) Infixing/suffixing alternation --- not applicable
 - Several Aux verbs (172ff.); but no bound affixes
- 8) Definite article in genitive embeddings --- not applicable
 - No articles
- 9) Nonconcord of V with full-NP Subj --- not applicable
- 10) Verbal abstract: VN or Inf? --- not applicable
 - No verbal abstracts per se; serial verbs appear in prototypical EQUI contexts (607, called "pivotal" constructions)
- 11) Predicative particle? --- NO
 - With PredNoun, use explicit copula: Subj Cop PredNoun (147ff.)
 - With PredAdj, Adj functions as V: Subj AdjVerb (142)
- 12) Prepositional periphrastic (Prep + VN) --- YES
 - Durative aspect (217ff.) coded either by zài (= Prep "at") preceding the Verb, or by -zhe suffixed to Verb; grammar does not identify zài as identical to the Prep, but they're at least homophonous, and written with the same character [WH, p.c.]
 - Only activity-verbs can take zài (218):
Zhangsan zài hit Lisi "Zhangsan is hitting Lisi"
 - Verbs expressing posture (219), or activity-verbs expressing associated state (221), take -zhe; and even activity-verbs take -zhe when used to express "durative background" (223), e.g.
she cry-zhe run return home go "She ran home crying"
- 13) "DO" periphrastic ("DO" + VN) --- NO [WH, p.c.]
- 14) Adverbial clause = "and" + finite clause --- not applicable
 - No "and" word (631ff.); no finite/nonfinite distinction
- 15) VN/Inf instead of finite main-clause form --- not applicable
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- barely

--- Many nouns (a finite closed set) must occur compounded with suffix zi, originally deriving from zi "child, seed" (now tonally distinct) (42-43); these are synchronically opaque; also "copper-zi-DIM = coin, "chess-zi-DIM" = chesspiece, "gun-zi-DIM" = bullet [WH, p.c.], all three semantically like "seed" (little round thing)

--- Also note compounds [WH, from dictionary]: "separate-zi" = numerator (of fraction) or molecule, "mother-gold" = capital, "child-gold" = interest, "mother-law" = constitution (law code), "child-law" = common laws, "character-mother" = alphabet, "dote-child" = favorite one, "liquor-mother" = liquid for fermenting grain or fruit; some of these may be borrowings (from Sanskrit or Japanese); semantics are a bit off for the desired construction

Chrau (Vietnam; Mon-Khmer [Austroasiatic])
Thomas 1971

0) Basic categorial information: Isolating (20, 152)

--- Noun: NP slot sequence (127); no affixes; no article

--- Verb: VP slot sequence (143); no person/number inflection

--- No article; the multi-purpose particle cô (see below passim) is described as article-like when used prenominally (85), and can cooccur with postnominal demonstrative or particle of surprise or emphasis (88-89); but will not reanalyze as "hedged article"

--- List of grammatical particles given (223-29)

--- No tones (19, but cf. 56-61 on "inherent word raising/lowering")

1) Conjugated adpositions --- NO

2) Word order

--- SVO (78)

--- Prepositions (98-99; also 83-84, 90-91); postpositional dative/benef element (70), identical to verb "give, have" (146)

--- N-Adj (139); but "Adj" really a subtype of verb, "Verbal Adj" (109-10, 148)

--- N-Gen (139)

--- N-RC1 (85, 105-126 passim)

3) Relative clause linker --- INVARIANT cô

--- Multiple uses of cô (84-86); prenominally, used for already-identified or secondarily focused referents; also copular use (see [11] below)

4) Relativization strategy/ies --- GAPPING

--- Syntagm: HeadN cô RC1 (84-85, 105ff., 174); RC1 must have explicit HeadN, very often pronominal ("he who...")

--- Grammar sometimes seems to treat RC1 as internally headed (174), but no argumentation; and one ex., lit. "(the) bamboo that I know he chops" (122), excludes any such analysis

--- On externally headed analysis (adopted here), coref N is gapped

--- Exx. and discussion (105ff.) mention only Subj and Obj RC1; list of transforms of basic clause-types fails to show a RC1 version under "Bitransitive" (118) and "Possessor Adjective" (111) clause-types, unlike basic intrans and trans clause-types (108, 112), which do show a RC1 transform; Obj of dative/benef "Postp" normally not deletable (71); hence will assume genitival and prep RC1 nonexistent

--- Another construction: HeadN may be followed asyndetically by a "reduced clause" (139), e.g. "(a) hen sit(ting) (on) eggs"

- 5) Special relative form of verb --- NO
- 6) Polypersonal verb --- zero actants coded
- 7) Infixing/suffixing alternation --- not applicable
- 8) Definite article in genitive embeddings --- not applicable
 --- No article; genitive syntagm is: HeadN DeptN (139)
- 9) Nonconcord of V with full-NP Subj --- not applicable
 --- No mention of noun plural; pronouns have optional pluralizer (138), and classifier for "person" has distinct suppletive sg and pl forms (131)
- 10) Verbal abstract: VN or Inf? --- not applicable
- 11) Predicative particle?
 --- Syntagm: Subj (Cop) PredN (Cop borrowed from Vietnamese)
 also: Subj cô[^] PredN (75-76)
 where cô[^] also serves as Rel Particle and quasi-Article
- 12) Prepositional periphrastic (Prep + VN) --- no mention, assume NO
 --- None of the Preps listed (98) recurs as preverbal particle etc. (143ff.)
- 13) "DO" periphrastic ("DO" + VN) --- no mention, assume NO
 --- No special mention of "do" under "compound verbs" (149-51); one compound verb mentioned of form "DO+Noun" (151)
- 14) Adverbial clause = "and" + finite clause --- not applicable
 --- No clause-level "and"-word mentioned under discussion of sentence conjunctions (171-74); most conjunctions borrowed from Vietnamese anyway (173); NP-level "and" word (140-41)
 --- Asyndetic clause coordination (176)
- 15) VN/Inf instead of finite main-clause form --- not applicable
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- no data

Chumash (Ineseño) (California; Hoka)
Applegate 1972

- 0) Basic categorial information
 - Noun: no cases; no gender; optional marking of plural; prefixed article ma-/ha- (221); possessive prefixes, identical to V subject markers (166, 235)
 - Verb: Subj and Obj person/number marked affixally (166, 167)
- 1) Conjugated adpositions --- no category "adposition" (431)
- 2) Word order
 - VOS (470, 476)
 - N-Gen (246)
 - N-RC1 (508), occasionally RC1-N; if RC1 expresses an adjectival-stative verb, either order equally common (509); explicit head N quite rare (508)
 - No adpositions; no adjectives (expressed by stative verbs, 509)
- 3) Relative clause linker --- invariant particle
 - Preverbal particle ma-/ha- (ma- initially, ha- medially), functioning as nominalizer (204, 309, 505); functions as article when preceding nouns; also relative particle ka (506-7)
- 4) Relativization strategy/ies
 - Verbs are nominalized (usually with prefixed ma-/ha-), but verb is otherwise normal finite verb (except for changes noted below, [5])
 - Headless relatives thus the norm (508); for intransitives, RC1 refers to agent of V (205); for transitives, almost always refers to object (205); verb can take locative suffix -pi, in which case RC1 refers to location (310)
 - Gapping for object and locative RC1 (inferred from exx.)
 - Copying for genitive RC1 (509, 516, see exx.); no prepositional RC1
- 5) Special relative form of verb --- YES
 - 3rd-person verb-subject prefixes assume special relative forms (e.g. 3sg s- becomes 3sg.rel l- after ma-); 1st and 2nd person markers stay the same (303)
- 6) Polypersonal verb --- 2 affixal arguments, Subj and Obj (166, 167)
- 7) Infixing/suffixing alternation --- NO
 - Order of affixes is always: Subj-V-Obj (166, 167)
- 8) Definite article in genitive embeddings
 - Article ma-/ha- precedes most common nouns in most syntactic environments, but has no connotation of definiteness or previous mention (221)
 - Syntagm (X of Y): Art-his-X Art-Y (article on both, per exx.)
Art-his-clan Art-eagle = eagle clan (246)
- 9) Nonconcord of V with full-NP Subj --- sort of

--- Verb: Subj coded prefixally (21, 101, 105), though a few verbs have suffixal Subj coding (83, 130); pronominal Obj coded suffixally (40) or as independent word (i.e. as Acc Prep n + Obj, 35); Subj marker typically preceded by a tense/asp/mood prefix, thus:

Pfx-Subj-VerbStem-Obj ;

a full NP can be incorporated into verb in place of Subj marker (22)

--- Article: prefixed, both Def (2) and Indef (5; note plural Indef form):

m-sg	Def: p(e)-	Indef: ou-
f-sg	t(e)-	"
pl	n(e)-	hen-

1) Conjugated adpositions --- YES (30)

2) Word order

--- SVO or VSO (22, 123); three possibilities for Subj marking:

(a) VS:	Pfx-PronSubj-VerbStem	NounSubj
(b) SV:	NounSubj	Pfx-PronSubj-VerbStem
(c) "SV":	Pfx-NounSubj-VerbStem	(noun incorporation)

--- Prepositions (3, 30); all are proclitics (3)

--- N-Adj or Adj-N (57)

--- N-Gen (6)

--- N-RC1 (8, 43)

3) Relative clause linker --- INVARIANT (8, 105-7)

--- Allomorphs et, ete, ent; choice not dependent on role of HeadN in RC1, but on tense and on verbal vs. nominal nature of embedded clause; different marker e- for RC1 on indef head; see [5]

4) Relativization strategy/ies --- COPYING (43)

--- All non-Subj RC1s (including DObj) require resumptive pronoun; with Subj RC1s, coref NP always deleted, while PronSubj marker in verb is retained in some tenses (43, 86), deleted in others (76)

--- RC1 may be headless, with DefArt p(e)- etc. in place of HeadN (44)

--- RC1 verbs presented in grammar as constituting a special "relative" tense series embodying prefix et- (see [5]); these are usable only with definite HeadN (8)

--- With indef HeadN, RC1 instead uses circumstantial verb-form (prefix e-), normally an adverbial ("while VERBing") (96, 107)

5) Special relative form of verb --- NO

--- Depends on analysis; verbs assume "relative forms" in et-, but transparently segmentable (101, 105):

Non-rel: Pfx-PronSubj-Stem Rel: et(e)-Pfx-PronSubj-Stem
(in one tense, et- assumes allomorph ent-, with "intrusive n" characteristic of Sahidic); moreover, et also introduces verbless (nominal) RC1s (8); hence will analyze et- as RelPtcl

--- Similarly for indef Heads: Circumstantial marker e- can be taken as superadded to normal tense forms (101, 105), hence particle

- 6) Polypersonal verb --- 2 actants coded (normally)
 --- In Causative, 3 actants coded; verb encodes both causer(1) and causee(2) with Subj-pronoun morphology in different slots, using "inflected infinitive" (80, 137):
 Pfx - Subj1 - [tre-Subj2-Stem] - Obj
- 7) Infixing/suffixing alternation --- NO
- 8) Definite article in genitive embeddings
 --- Syntagm: the-X n the-Y (n = "of") (6)
- 9) Nonconcord of V with full-NP Subj --- in a sense
 --- In one Subj-Verb syntagm (see [2] above), full-NP Subj replaces PronSubj (22); here verb lacks any pronominal marking
- 10) Verbal abstract: VN or Inf? --- INF
 --- "Infinitive" is the basic verb-stem used in building tenses (see [2]), also used without prefixes to express purpose (49) or EQUI (80); there's also an "inflected infinitive" (80)
 --- Pronoun Obj either acts as indep NP (35) or as enclitic to verb (39-40), but not proclitic as with pronominal possessor (11); full-NP Obj may be marked with Acc Prep n (35), homophonous to genitive Prep n "of" (6), but may also be suffixed to proclitic "prenominal" form of verb (37); thus syntax is clearly non-genitival
- 11) Predicative particle? --- as an option
 --- Basic copular syntagm is: Subj pe/te/ne PredN/Adj, in various orders, with pe/te/ne identical to Def Art and functioning as copula (14, 59); PredAdj must take indefinite article (59), PredN may take either article
 --- Another option: "Qualitative" (stative) form of r- "do, become" is o "be", and of šōpe "occur, become" is šōop "be"; both verbs take predicative particle n (94, 111), homophonous to n "in"; Pred does not take article; ex.:
 o/šōop n ponēros "to be wicked"
- 12) Prepositional periphrastic (Prep + VN) --- NO
 --- "Inflected Inf" can occur in adverbial use following Preps meaning "in", "after" (81); but not with copula
- 13) "DO" periphrastic ("DO" + VN) --- in a sense
 --- Lots of compound verbs, many formed with r- "do, make" plus a nominal element (110-12); no exx. given where nominal element is itself a verb stem or infinitive, though it may be a deverbal nominal
- 14) Adverbial clause = "and" + finite clause --- NO
 --- Clause-level conjunction "and" = awō (27, 142), no mention of Adv-clause usage; cf. discussion of conjunctions (140-43)
- 15) VN/Inf instead of finite main-clause form --- NO
 --- Two different "and-then" tenses ("Conjunctive"), continuing the tense of previous verb (107-8, 135-36); but such clause-chaining forms are finite

16) Word-initial change --- NO

17) Extended use of kin terms ("Kin of Noun") --- NO

--- Only one ex. noted in Crum dictionary: "daughter of eyes" = pupil, calqued on Greek (where kóre means "daughter" and "pupil")

Cree (Canadian Great Plains; Algonquian)
Wolfart 1973; Wolfart & Carroll 1981, Ahenakew 1987
[My thanks to Rich Rhodes for indispensable help throughout]

0) Basic categorial information

--- Noun: animate/inanimate gender distinction (20-23); distinction of proximate (NPs which are "in focus" or "center of attention") vs. obviative (other NPs) (16-17); plural coded suffixally (23, 28), conflated with prox/obv; no case, but do have a locative suffix (31), mutually exclusive with Numb/Obv suffix; pronominal possessor coded by circumfix (prefix showing person, suffix conflating person/number) (15-16); slot sequence:

PossPfx - N - PossSfx - {Loc, Numb/Obv} (28, 31, 1981:50-51)

--- Verb: intransitive verbs code Subj very much like pronominal possession; transitive verbs code both Subj and Obj, according to complex "Direct/Inverse" system (24-26): with Direct forms, prefix shows person of Subj, suffixes (fairly well separable) indicate pers/numb of Subj and of Obj; with Inverse forms, all morphs referring to Subj (in Direct form) now refer instead to Obj, and vice versa; ditransitive verbs code notional IObj (not DObj) (75); also, distinct from the "Independent" inflection just described, have a "Conjunct" inflection, purely suffixal and involving a different set of endings (1987:48)

--- Article: none mentioned; but demonstratives (which may function either as pronouns or modifiers, 33) are very frequent in texts, can behave very much like articles [RR, p.c.]; will analyze as "hedged article"

1) Conjugated adpositions --- probably irrelevant

--- Only two free postpositions, isi "to" (manner/goal) and ohci "from" (source, reason) [RR, p.c.] (cf. p. 77); these may occur as free forms, but tend strongly to be incorporated into the verb (which incorporates a good many such Postp-like elements, most not occurring as free forms); Postps can only have 3-person Objects, and these mostly non-referential [RR, p.c.]; no information about pronominal Objects

2) Word order [RR, p.c.]

--- V-first (VOS unmarked order); but NP conveying new information occurs preverbally

--- Postpositions (but the category barely exists)

--- No category "Adj"; verbs instead (1981:85)

--- Gen-N or N-Gen (exx. 25, 29, 1981:42); but see [8] for dubious status of dependent genitive

--- RC1 basically headless, possibly internally headed (see [4] below)

- 3) Relative clause linker --- irrelevant (see [4])
- 4) Relativization strategy/ies
- RCl basically headless ("he who..."); verb must appear in Conjunct form (characteristic of subordination in general), in its "participial use" (46); no formal mark within RCl indicating which participant counts as conceptual Head [RR, p.c.], may be Subj or Obj or obliques ("when, where")
 - A full-NP "HeadN" sometimes appears; exx. in grammar books show "HeadN" both before and after RCl (1981:36, 1981:74); two conceivable analyses:
 - (a) RCl appositional to HeadN
 - (b) More likely, RCl internally headed [RR, p.c., per work by Amy Dahlstrom]; most exx. amenable to either analysis; must concentrate on RCls of the form, e.g.:

Verb-Conjunct	HeadN	X	"HeadN who Verbs X"
---------------	-------	---	---------------------

 with X belonging to subordinate clause, for here HeadN is "boxed in"
 - "Postpositional" RCls will involve incorporation of Postp within verb [RR, p.c.]; but category "Postp" marginal
 - Genitival RCls: unclear, no exx. come to mind [RR, p.c.]
- 5) Special relative form of verb --- not exactly
- Verb must appear in Conjunct form (46), and requires the initial morpheme called "The Change" [RR, p.c.]; but these forms may appear in other subordinate contexts as well
- 6) Polypersonal verb --- two actants coded
- 7) Infixing/suffixing alternation --- weakly
- For any given pers/numb specification of Subj and Obj, slots of Subj and Obj are fixed
 - But technically do have a partial prefixing/suffixing alternation regarding Indep vs. Conjunct forms: Conjunct forms have only suffixal coding, while Indep forms also involve prefixes; in Direct (Indep) forms, prefix indicates Subject, while in Inverse (Indep) forms, prefix indicates Object; hence for Inverse forms, Obj coding alternates: prefixal in Indep, suffixal in Conjunct
 - Conditioning factor is not preverb
 - Argument seems strained [RR, p.c.]
- 8) Definite article in genitive embeddings --- hedged
- Genitive seems fundamentally a matter of a possessive-marked HeadN ("his-book"), which may optionally have a full-NP appositional to "his", either preceding or following; not a real "dependent genitive"
 - Demonstrative-qua-article may accompany a possessed noun ("his-book this"), and can appear on both NPs in "Headed" genitive syntagm [RR, p.c.]
- 9) Nonconcord of V with full-NP Subj
- No sg/pl distinction for Animate Obviative in any grammatical category (14, 24); instances where notionally plural Subj is expressed as singular (collective), on both Noun and Verb (24)
 - The only sg/pl mismatch involves instances of "split reference", e.g.:

My-wife we-went [= my wife and I] to town. [RR, p.c.]

10) Verbal abstract: VN or Inf? --- NEITHER

--- All verbs conform to standard person/number coding schemas (either Indep or Conjunct) for Subj, Obj; subordination involves fully finite Conjunct forms, with Subj and Obj both coded (46)

11) Predicative particle? --- NO [RR, p.c.]

--- Cree has distinct equational and locational copulas, both ordinary verbs; 3 possible equational constructions, none with particle:

- (a) Incorporate PredN into copular verb -iwi
- (b) Separate copula: PredN Cop SubjN
- (c) Omitted copula: PredN SubjN

12) Postpositional periphrastic (VN + Postp) --- not applicable

--- No verbal abstracts, barely any Postps

13) "DO" periphrastic ("DO" + VN) --- not applicable

--- No such use of "do" [RR, p.c.]

14) Adverbial clause = "and" + finite clause --- NO [RR, p.c.]

--- There is an "and"-word ēkwah, functioning at both NP-level (ex.

14) and clause-level (ex. 15); but no adverbial use

--- Adv clauses use Conjunct verb forms

15) VN/Inf instead of finite main-clause form --- not applicable

--- Conjunct forms can function in clause chaining, occurring in all but the initial verb [RR, p.c.] (46, 1987:157); but these are finite forms

16) Word-initial change --- not really [RR, p.c.]

--- In certain syntactic/semantic contexts, Conjunct verbs appear in a form called "Changed"; this "Change" involves an alternation in the first vowel of the verb (whether or not in absolute initial position), but only if that vowel belongs to a restricted set of morphemes (under a dozen); otherwise the "Change" is realized not as vowel alternation but as a prefixed e-

--- "Change" is also extensively involved in derivational processes

17) Extended use of kin terms ("Kin of Noun") --- NO [RR, p.c.]

--- Note: Genitive construction requires an animate possessor, thus incompatible with exx. like "son of the road", etc.

Dyirbal (Australia; Pama-Nyungan)
Dixon 1972

- 0) Basic categorial information: Ergative/Absolutive
- Noun: inflects suffixally for case (42, 236ff.), with possibility of possessor N being doubly case-marked (genitive + case of HeadN) (43); plural marked optionally, usually by reduplication, sometimes suffixally (241-42); noun class indicated by prenominal "noun marker" (44) [see "Article"]; nouns case-marked ergatively (with "Absolutive" case called "nominative"), but Pronouns marked Nom-Acc (59-60); syntax largely ergative
 - Verb: no actants marked on V (54ff., 246ff.)
 - Article: none described, but "noun marker" can plausibly be taken as Article: it's a normal (though not universal) concomitant of noun (60), agrees with it in case and noun class, also encodes proximity of referent (3 degrees), hence determiner-like (44-45)
- 1) Conjugated adpositions --- not applicable
- No category "adposition"
- 2) Word order: essentially free (291), discontinuous constituents very possible (ex. 107); the following are merely preferences
- OSV (= Nom Erg V) (291)
 - No category "adposition"
 - N-Adj (60, 291)
 - Gen-N (291)
 - N-RC1 (103, 291)
- 3) Relative clause linker --- ZERO
- 4) Relativization strategy/ies --- GAPPING
- RC1 deletes coreferential N; Verb replaces tense-suffix by special RC1 ending -gu + Case, where Case agrees with case of Head N (99); Mamu dialect additionally has similar RC1 ending -mi, with same syntax (103)
 - Coref N in RC1 must be in nominative case (100); antipassivization with -pay allows relativization on Ergative Subj; this constraint would seem to preclude genitival RC1s
 - Also various "participial" forms (81ff.), function adjectivally:
 - intrans: one who Verbs
 - trans: one who is Verbed
- 5) Special relative form of verb --- YES
- Endings -gu, -mi as verb endings occur only on RC1; on nouns, function as 2 different marks of Genitive; Dixon thus analyzes Genitive as reduced RC1 (176ff.)
 - But as verb endings, they uniquely mark V as a RC1-form
- 6) Polypersonal verb --- zero actants
- 7) Infixing/suffixing alternation --- not applicable
- 8) Definite article in genitive embeddings
- Three constructions (42-43, 61-62, 105-10):

- (a) Apposition, for inalienable possession
 - (b) "Simple genitive" case, for alienable possession: either
 - (i) DeptN-smp.gen HeadN (nominative HeadN)
 - (ii) DeptN-smp.gen-din-case HeadN-case (non-nominative)
 - (c) "General genitive" case, for alienable possession in past: either
 - (i) DeptN-gnl.gen HeadN (nominative HeadN)
 - (ii) DeptN-gnl.gen-case HeadN-case (non-nominative; no -din-)
- NB: DeptN can take second case marker, agreeing with case of HeadN; order of HeadN and DeptN flexible
- Will here take "noun marker" as Article; N-marker has only a "Simple genitive" form (44), used in either genitive case (108-9)
- Grammar has exx. where N-marker occurs only on DeptN (inalienable possession, 61; General genitive, 108), only on HeadN (General genitive, 109), and on both HeadN and DeptN (Simple genitive, 105)
- 9) Nonconcord of V with full-NP Subj --- not applicable
- 10) Verbal abstract: VN or Inf? --- INFINITIVE (?)
- Under "Nominalizations" (81ff.), grammar mentions only participles, not verbal abstracts
- Since V lacks person/number coding, only lack of tense can specify a given form as "verbal abstract"
- "Purposive" inflection (67ff.) replaces Tense by Purp ending; productive process for any verb (69); semantics roughly appropriate to purposive EQUI (exx. passim), also occurs after "tell" (75)
- Structurally, Purposive usually occurs on noninitial V in topic-chaining; with transitive Purp verb, notional DObj will occur in Nom (as usual) or in Dat (if antipassive), but certainly not in Gen; technically this counts as "infinitive" (since Obj is in same case as in finite clause)
- 11) Predicative particle? --- NO
- No explicit discussion of predicate N or Adj; exx. show NounPred (46, 210-11) in Nom case, and AdjPred (205) apparently inflecting for tense; but no particle
- 12) Adpositional periphrastic (Adp + VN) --- not applicable
- 13) "DO" periphrastic ("DO" + VN) --- no mention, assume NO
- 14) Adverbial clause = "and" + finite clause --- not applicable
- No discussion of conjunctions, no sign of an "and"-word
- 15) VN/Inf instead of finite main-clause form --- marginally
- The only clear candidate (?) for "VN/Inf" is Purposive, which usually occurs as noninitial verb in clause-chaining (68), but can also occur first (relatively infrequently, 68-69) if an "implicating" event has occurred but not been mentioned
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- no data

English (Britain; Germanic [Indo-European])

- 0) Basic categorial information
 - Noun: gender (masc/fem) coded only on personal pronouns, not nouns; 3-way case distinction (Subj/Obj/Gen) on pers pronouns, 2-way distinction on nouns (General/Genitive ['s]); noun plural coded suffixally
 - Verb: two synthetic tenses (Pres, Past), rich periphrastic tense system for future, modals, progressive, and perfect (with preverbal Aux); pers/numb not marked in Past, minimally in Pres (3sg vs. other); only the verb "BE" has relatively full pers/numb marking
 - Article: both definite and indefinite, pronominal; either proclitic or stressed independent word
- 1) Conjugated adpositions --- NO
- 2) Word order
 - SVO
 - Prepositions
 - Adj-N; but usually N-AdjPhr for larger AdjPhr
 - Gen-N (with 's'), or N-Gen (with of)
 - N-RCI
- 3) Relative clause linker --- three types
 - Either zero, or invariant relative particle (that), or relative pronoun (which etc.); must use RelPron (whose) in genitive RCI; cannot use zero in Subj RCI
 - RelPron (initial in RCI) distinguishes animate (who) and inanimate (which); animate distinguishes Subjective (who) vs. Objective (whom) case; both have special genitive form (whose)
- 4) Relativization strategy --- GAPPING
 - Coreferential NP gaps under identity; RelPron (if this linker type is selected) appears clause-initially as pronominal copy
 - Genitive RCI: must use RelPron; entire possessed NP fronts, with coref possessor converted to whose, leaving gap in situ
 - Prep RCI: two strategies:
 - (a) Front entire PP to clause-initial position, with coref NP converted to RelPron, leaving gap in situ
 - OR (b) Gap coref NP, but strand Prep in situ; all 3 linkers possible
- 5) Special relative form of verb --- NO
- 6) Polypersonal verb --- 0 or 1 actant (Subj)
 - Pres codes 3sg, Past shows no coding; "BE" codes most pers/numb
- 7) Infixing/suffixing alternation --- not applicable
- 8) Definite article in genitive embeddings
 - N-Gen type with of: each NP takes its own Article
 - Gen-N type with 's': only Dept NP can take an article:
 - [a boy's] dreams ;

pattern holds equally for def and indef article, but ex. with indef article proves that article in fact goes with Dept (*a dreams)

9) Nonconcord of V with full-NP Subj --- NONE

10) Verbal abstract: VN or Inf? --- usually INFINITIVE

--- Infinitive with "to", pronominal object always in Objective case

--- Gerund in -ing also productive; rection usually accusative ("planting trees"), may also involve genitive ("the planting of trees")

11) Predicative particle? --- NO

--- Syntagm: Subj BE Pred

12) Prepositional periphrastic (Prep + VN) --- NO

--- Only Purpose usage exists, "He is to go"; many other periphrastic tenses, but none with Prep

13) "DO" periphrastic ("DO" + VN) --- YES

--- Syntagm: DO + VerbStem; this is the standard technique for forming questions and negatives, also for emphatic positive statement

14) Adverbial clause = "and" + finite clause --- NO

15) VN/Inf instead of finite main-clause form --- NO

16) Word-initial change --- NO

17) Extended use of kin terms ("Kin of Noun") --- NO

Eskimo (West Greenlandic) (Greenland; Eskimo-Aleut)

Fortescue 1984

[My thanks to Anthony Woodbury for answers to several queries]

0) Basic categorial information: Polysynthetic

--- Noun (205-9): ergative/absolutive case marking ("Absolutive" vs. "Relative" [ergative] case, the latter also used for possessor) (210); number is marked, also possessive suffixes; no articles (110), no gender (248)

--- Verb: inflects for Subj and Obj person/number (but not IndObj) (298); frequent noun incorporation, notably of Obj (82ff.)

1) Conjugated adpositions --- YES (but actually relational nouns)

--- No categorial distinction between "adposition" and "head N in possessive" (106-7); analyze here as "hedged postpositions"; possessed nouns always take possessive suffixes, even with full-NP possessor (216), hence adpositions do as well

2) Word order

--- SOV (93)

- Postpositions (or rather, relational nouns) (106-7)
 - No category "Adjective" (49, 204); adjectival function is filled by stative verbal and nominal bases (204), which follow the N (108)
 - Gen-N (109)
 - N-RC1 (51)
- 3) Relative clause linker --- ZERO
- 4) Relativization strategy/ies --- PARTICIPLES; GAP/COPY for possessor
- Use participles: for transitive verbs, must be passive participle; RC1 directly follows Head-N (51, 108); participle (final in its RC1) agrees with Head-N in person/number/case (52)
 - Can relativize (52-55) on embedded Intrans-Subj, Trans-Obj; to relativize on Trans-Subj, must intransitivize the verb (the "half-transitive" form, taking its object in Instrumental, 85-86; apparently a sort of antipassive [though not so characterized])
 - Possessor relativization possible but disfavored, involves gapping the Noun and leaving intact the pronominal copy (53); other oblique relations relativizable by using a derived verbal base (in effect, voice-changing)
- 5) Special relative form of verb --- not exactly
- "Participle" used not just for RC1s, but also for Object-complement clauses (34ff.) and adverbial clauses (56ff.); counts as a full "mood" of the verb (289-90), can take full person/number inflection
- 6) Polypersonal verb --- 2 actants (Subj, Obj) coded on verb (298)
- 7) Infixing/suffixing alternation --- NO
- Verbal Subj/Obj marker, though presented as sequence "Subj+Obj" (289), seems largely to conflate the categories (an analysis as "Subj+Obj" involves much nontransparent (?) allomorphy); apparently fixed position in V slot-sequence
- 8) Definite article in genitive embeddings --- not applicable (no articles)
- NB: Possessor and Possessed both marked affixally (Gen-N: "hunter-'s kayak-his") (216; see 206ff. for possessed paradigms)
- 9) Nonconcord of V with full-NP Subj --- NO (298)
- 10) Verbal abstract: VN or Inf? --- VN
- Only one V form is termed nonfinite: -niq (44ff., 297); lacking in person/number and mood marking; usable in most contexts where an Obj complement-clause is needed (48, cf. 35-44), though (apparently) the straightforward way of forming Obj complement-clauses is rather via participles or Contemporative mood; these Obj-complements include various EQUI contexts
 - Subj of -niq verb (if present) is in Relative case (45) (= the case used for Possessor or Trans. subject)
 - To express transitive Object, -niq verb may convert to "half-transitive" (45, 212-13), with Obj now in Instrumental (85); or, instead, Obj may itself appear in Relative case ("the killing of the man", 46), apparently (?) when there is no Subj expressed

--- In neither instance will Obj take Absolutive case (as in finite clauses), hence not Infinitive; and construction with Obj in Relative case should count specifically as "VN"

--- There is a "Contemporative" (Ctmpv) verb-mood (297), with various uses: Object complement-clauses (34ff.), Adverbial clauses (55), notional clause coordination (120ff.), clause chaining (usually Same-Subj, 299-300); grammar book presents it as "a true inflectional mood" (297), with (apparently) Obj marked as usual; appears to be a normal finite form, but with one exception: Subj is usually NOT marked at all on transitives (297); however, this will indeed count as finite because form does mark S or O, as is natural for an ergative language; note that Ctmpv occurs in at least one environment which is prototypical for "VN/Inf", viz. Obj complement-clause

11) Predicative particle?

--- Nouns can take a verbalizing suffix -u (70), meaning "be a N"; or can use equational "copular particle" tassa (72), linking the nouns (both usually in Absolutive)

12) Postpositional periphrastic (VN + Postp) --- no mention, assume NO

--- -nig nominalizations may be used with Postpos, but only to express adverbial (temporal) notions (62)

--- Recall that Postp is "hedged postposition"

13) "DO" periphrastic (VN + "DO") --- no mention, assume NO

14) Adverbial clause = "and" + finite clause --- apparently YES

--- A particular usage of Contemporative (a finite form, recall [10]) involves enclitic -lu "and" (120-22, also 59):

V1(Ctmpv)-lu V2(main) = "he V2ed as soon as he V1ed"

--- Also, Ctmpv and -lu can be used in an "absolutive" type of construction (55); example:

How will I ever become a shaman, uanga-lu nalu - llunga
I -and not.know-Ctmpv.1sg
"How will I ever become a shaman, me being ignorant?"

15) VN/Inf instead of finite main-clause form --- not really

--- True nominalization (-nig) not used (apparently) to express main predicates (44)

--- Contemporative mood appears in Same-Subj clause-chaining (297, 301-2); Ctmpv also has VN/Inf uses; but better taken as a finite form (see [10])

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- NO [AW, p.c.]

Fijian (Boumaa) (Pacific; Oceanic = Eastern Austronesian)
Dixon 1988

0) Basic categorial information

--- Noun: no cases; no gender, but there are classifiers, occurring only in possessive constructions (135); number not coded inflectionally (114); possessive suffixes (36, 121, paradigms 54-55)

--- Verb: Subj and Obj person/number not marked affixally, but by pronouns specialized for use with V (53-55); usually these pronouns obligatory even with full-NP arguments (35, 68)

--- Article: nouns usually preceded by "article" (common vs. proper), which (NB) is semantically "unmarked for definiteness" (35-36, 114-16)

1) Conjugated adpositions --- YES, 3-sg only

--- Only two or three Preps exist ("to, from, with") (151ff.); apparently take "Cardinal pronouns", which occur "when a pronoun occurs as head of an NP" (53); but 3-sg (and only 3-sg) has special fused forms Prep+Pron (57, 151); quite distinct from possessive suffixes on nouns (54-55, 121)

--- When Obj-of-Prep is a human full-NP (and optionally for nonhuman), proleptic pronominal Obj intervenes ("to them the people", 114, 152); includes the fused 3sg cases

2) Word order

--- VOS (35)

--- Prepositions (few) (151ff.)

--- N-Adj (36, 117)

--- N-Gen (36, 119ff.)

--- N-RC1 (36, 251)

3) Relative clause linker --- ZERO (252)

--- But Old High Fijian does often use conjunction ka "and" to introduced RCls (252, 259)

4) Relativization strategy/ies

--- "There is complete absence of any formal mark of a RC1" (252); just the ordinary clause, minus coreferential NP

--- If relativized NP is in "peripheral" (= oblique) role in RC1, then predicate of RC1 must contain resumptive element ?ina (251, 255), analyzable as fusion of Prep i + Pron na "to/at it" (57, 63, 248); i.e., Copying; this ?ina does not remain in situ but moves forward in RC1, to shortly after the predicate head (63, 248)

--- ?ina similarly usable as resumptive for NPs that are fronted (preverbal) in nonrelative clauses (248)

--- ?ina not required if NP serves as Subj or Obj in RC1 (255); NP gaps, pronominal copy on verb remains

--- Relativized noun can be "the possessor in an NP" (255), but no details given as to construction

5) Special relative form of verb --- NO

6) Polypersonal verb --- usually zero actants (sometimes Subj proclitic)

--- Special forms for Subj (preverbal) and Obj (postverbal) pronoun as

occurring in the "predicate complex" (53-55); Obj pronouns are always, and Subj usually, independent phonological words (30); 3-sg subj e (often omissible) is proclitic (to Verb), several other Subj pronouns are enclitic to other preverbal particles (30)

--- These pronouns are integral parts of the predicate complex (64, 68); Cardinal (independent) pronouns may cooccur with them

7) Infixing/suffixing alternation --- NO

--- Pronoun positioning within predicate complex is fixed: Subj V Obj

8) Definite article in genitive embeddings

--- NPs usually begin with articles (35); there is a proper article and a common article, the latter unmarked for definiteness (114); calling these by the name "article" is established usage in Fijian grammar (114)

--- 5 distinct N-N possessive constructions (35, 120-22, 124-26), some using Classifiers (Clsf):

- | | | |
|-------|---|-------------------------|
| (i) | Art Head- <u>i</u> Poss | (Poss = proper name) |
| | --- "the hand- <u>i</u> John" | |
| (ii) | Art Head _i Clsf _i - <u>i</u> Poss | (Poss = proper name) |
| | --- "the _i bread _i Clsf- <u>i</u> John" | |
| (iii) | Art Head-Suffix [Art Poss] | (Poss = human noun) |
| | --- "the hand-her [the girl this]" | |
| (iv) | Art Clsf _i -Suffix Head _i [Art Poss] | (Poss = human noun) |
| | --- "the _i Clsf-their boat [the Mika and John]" = "M & J's boat" | |
| (v) | (Art) Head <u>ni</u> Poss | (Poss = non-human noun) |
| | --- "the teeth <u>ni</u> horse" | |

--- In (i-ii), Poss apparently may not take the article (123); in (iii-iv), Poss is an appositional separate NP (with Art), not in direct construction with Head (121-22); in (v), Poss may not take article following ni (115, 124)

--- In pronominal possession, Head takes suffix-pronouns and also Article

--- Thus Head always must or may take article

9) Nonconcord of V with full-NP Subj --- minimally

--- Nouns have no plural mark; plurality indicated by apposition to plural pronouns (114) ("they our-ancestor" = "our ancestors"); no indication that verb concord is suspended here

--- When full-NP subject includes a numeral (or kindred word), then verb may optionally be 3-sg if Subj is inanimate, but must show number concord if Subj animate (146); this concord is realized on the "subject pronoun" which is part of the predicate complex

--- Verbs must show number concord with conjoined Subj NPs (157); syntagm actually "N1 with N2", where N1 is deletable (157ff.)

10) Verbal abstract: VN or Inf? --- NEITHER

--- Any clause can be nominalized ("Clausal NP", like English Poss-ING: "his watching"); clause is preceded by article, Subj is realized as Possessor, otherwise all other components of predicate remain unchanged (130, 132), in particular, Obj continues to be marked as such; treat this as finite clause

--- Subj pronoun normally cannot be deleted even under coreference or EQUI (68, 259)

--- Verb complement: either Clausal NP, or full clause introduced by conjunctions (called "Relators") (267-68); me is Relator used especially for "to"-clauses (268)

11) Predicative particle? --- NO

--- Adj or N predicates occur initially, no particle (64-66); also "equational" sentence (N N) with "predicate" as second NP (again no particle) (241)

12) Prepositional periphrastic (Prep + VN) --- not applicable

--- Fijian has no copula, no VN

13) "DO" periphrastic ("DO" + VN) --- not applicable

--- No VN; no mention of periphrasis with verb "do"; see pp. 279ff. for "semi-auxiliary" verbs, which do not include "do"

14) Adverbial clause = "and" + finite clause --- not applicable (?)

--- No mention, probably not applicable; see pp. 258-59 for conjunctions ("Relators"), etc.; apparently no "and" conjunction

15) VN/Inf instead of finite main-clause form --- not applicable

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- a tiny bit

--- "mother of web" = spider; "child of lost" = orphan, abject person; "father" used for very large things ("father of canoe" = very large canoe); "child of water" = secret society believing in water-elves (Capell 1941)

French, modern spoken (Western Europe; Romance [Indo-European])

[My thanks to Jean-Pierre Koenig for checking my French intuitions]

0) Basic categorial information

--- Noun: categories are gender and number, coded indirectly through proclitic articles and concord patterns; some nouns have distinct plural forms in all contexts, and most have distinct plurals (suffix -s) when preceding vowel-initial words in "liaison" contexts

--- Verb: Subj, DObj, IObj encoded usually proclitically, but enclitically on Imperatives; true suffixal endings for 1-pl and 2-pl, and some verbs preserve a fuller person/number inflection (suffixal); full-NP arguments typically doubled on the verb as clitics; contrast written French, where the clitics do not count as "verb inflection" and where a full person/number paradigm exists in the spelling; also periphrastic tenses (perfect)

--- Article: definite (coding gender/number) and indefinite (coding gender, singular only), proclitic to NP

- 1) Conjugated adpositions --- NO
 - Prepositions, noninflecting (PrepObj pronouns belong to "emphatic" series)
- 2) Word order
 - VSO, SVO (vs. SVO in written French)
 - Prepositions
 - N-Adj, with some adjectives standardly Adj-N; order often reversible for stylistic effect
 - N-Gen
 - N-RCI
- 3) Relative clause linker --- Relative pronoun
 - Rel Pron (initial in RCI) codes case within RCI: distinct forms for Subj (qui), DObj (que), Gen (dont, duquel); Prep relatives (including IObj) marked with [Prep + lequel/qui] (analogous to English "with which", etc.); forms qui/que/dont invariant, forms with -quel code gender/number of HeadN
- 4) Relativization strategy/ies --- GAPPING with Rel Pron
 - Any role except Obj-of-Comp relativizable; NP is gapped, with case role reflected in choice of RCI-initial Rel Pronoun; for genitive RCI, gap both NP and Prep "of"; for Prep RCI, gap NP and move Prep to beginning of RCI along with Rel Pron: [Prep + RelPron]
- 5) Special relative form of verb --- NO
- 6) Polypersonal verb --- 3+ actants coded
 - Clitic complex (usually proclitic, enclitic in Imperatives) codes pers/numb of Subj, DObj, IObj, also includes special "to-it" (y) and "of-it" (en) markers
 - Pers/numb of Subj also coded suffixally (remnant of old inflection system) for 1pl, 2pl
- 7) Infixing/suffixing alternation --- YES
 - Non-subject clitics are suffixed to Imperative verbs; with non-Imperative verbs, occur infixed between Subj proclitic and verb stem
- 8) Definite article in genitive embeddings
 - Genitive uses "of" preposition de; syntagm is
 Head de Possessor
 - Each NP takes articles separately, exactly as in non-genitive contexts
- 9) Nonconcord of V with full-NP Subj --- NONE
 - The type "Il est venu des hommes" ("Some men came") is only apparently an exception: the verb est is singular, but it's an impersonal construction (as if "There/It came some men") [JK, p.c.]
- 10) Verbal abstract: VN or Inf? --- INFINITIVE
 - Only one verbal abstract, with straightforward finite rection
- 11) Predicative particle? --- NO
 - Construction is with copula, involves no particle; order is

BE Pred Subj (sometimes Subj BE Pred)

- 12) Prepositional periphrastic (Prep + VN) --- not exactly
 --- Progressive aspect usually not marked; can be coded using complex particle en-train-de (lit. "in course of"):
 BE en-train-de Inf
 --- As for other Prep+Inf, have only
 BE to Inf "(Subj) is to be Verbed" (Obligation)
- 13) "DO" periphrastic ("DO" + VN) --- NO
 --- "DO" + Inf has standard meaning of Causative; there are scattered lexicalized usages with "faire" (DO), such as "faire dodo" = "sleep", but a very minor pattern
- 14) Adverbial clause = "and" + finite clause --- apparently not
- 15) VN/Inf instead of finite main-clause form --- not in spoken French
 --- In literary narration, can use de+Inf to express equivalent of finite predications (see Grevisse, Le bon usage):
 Et les enfants de chanter maintenant "and now the children sing"
 --- Not clause chaining; archaic in modern speech
- 16) Word-initial change --- NO
- 17) Extended use of kin terms ("Kin of Noun") --- NO [JK, p.c.]

Gbeya (Central African Republic; Adamawa-Eastern [Niger-Congo])
 Samarin 1966

- 0) Basic categorial information
 --- Noun: no gender or case; plural optionally marked by preposed "plural word" (81); determinant and postclitic "articles" (see Article); pronominal possessor may be marked by possessive suffix on noun (102, 104), but only for "intimate nouns" (a kind of inalienable possession) (98); the pronominal suffixes exist only for 1sg, 3sg, 2pl, and come in two sets (104-5): subjectival (see Verb) and non-subjectival, the latter used as noun possessor, Obj-of-Prep, Obj-of-Verb (103); suffixal pronouns apparently not obligatory, may be free pronouns instead; see table of pronouns (101)
 --- Verb: no affixal pers/numb marking; six verbs have suppletive sg/pl forms, reflecting number of Subj for intrans verbs, number of Obj for trans (114); pronominal Obj (1sg, 3sg, 2pl) can be marked suffixally (103, 104); Subj may be Noun, free Pronoun, zero, or subjectival pronominal enclitic (1sg, 3sg, 2pl), the last attached to previous word (105, 129)
 --- Article: two candidates for "definite article":
 (i) Determinant suffix -a(a), basically deictic (48)
 (ii) Referential postclitic -i, basically anaphoric (52-53);
 both can cooccur (48); postclitic commoner in texts (52)

--- Analytical problem: allomorphy of 3sg pronominal suffix (104) and of Determinant suffix (48) is very similar, confusion possible

1) Conjugated adpositions --- somewhat

--- Only five "real" prepositions (73-75), all requiring an Obj except instr/loc nE; also, certain nouns used as pseudoprepositions (126, 117 n. 16, 142 n. 3)

--- Preps may take pronominal Obj as suffix (1sg, 3sg, 2pl only) or as free form (all pers/numb) (103-4); exx. passim of pseudopreps + suffixal Obj (e.g. 178, sentence 10: "underneath it")

2) Word order

--- SVO (125, 128-29); Adverbials after Obj

--- Prepositions (73)

--- Adj-N (80); two adjectives (with abstract semantics) have N-Adj (83)

--- N-Gen; two strategies, with (98, 74) and without (47) Prep "of"

--- N-RC1 (65)

3) Relative clause linker --- INVARIANT (= particle nE "and", 65)

4) Relativization strategy/ies

--- Syntagm: HeadN nE RC1 (65); referential postclitic -i "the" may appear twice, on HeadN and at end of RC1 (53; see [0] above)

--- For Subj RC1: if have head Noun, RC1 verb takes no Subj marker (123); if head Pronoun, Subj marker (= repeated Pron) apparently does occur (per exx. 65, 124)

--- Grammar states only that role of coref N in RC1 may be Subj, Obj, or Time/Place Noun (65); but I also found a few exx. of genitival and prep RC1:

Prep RC1 (nE): Gap (can't tell if Prep moves) (65 = 212)

Prep RC1 (nE): Gap/Move Prep (135; cf. 160 text 3:81, not true RC1)

ex.: pot [REL they prepare with ___ beer]-the (135)

"the pot with which they prepare the beer"

Prep RC1 (not nE): Copy (166 text 5:35)

PseudoPrep RC1: Copy (178 text 11:10; 142n.8 = 180 text 11:44)

Genitival (?) RC1: Copy (62 = 168 text 7:7) (analysis unclear)

--- Conclusion: genitival, prep, and pseudoprep RC1 all attested, use copying; Loc RC1, too, seems to take resumptive "there" (exx. 65, 123)

--- But Prep nE "with, about, at" has gapping instead, with Prep also (apparently) moving to verb from normal post-Obj position (per VP positional slots, 125); note that nE is the only Prep which need not take an Obj (75)

--- A different strategy: can use high-tone perfective and imperfective verb-forms (82, 113) (see [10] below), which can function adnominally rather like English participles

5) Special relative form of verb --- NO

6) Polypersonal verb --- maximum one actant coded, apparently

--- Pronom Obj can be suffix; Pronom subj enclitic to previous word (see [0] above)

--- "Previous word" may itself be a transitive verb (chaining), hence theoretically a verb may code two participants (= own Obj, anticipatory Subj); but no discussion or exx.; will ignore

7) Infixing/suffixing alternation --- NO

8) Definite article in genitive embeddings

--- Two "articles" (Determinant and referential postclitic)

--- Two genitive constructions:

(i) With purely tonal "relational affix" (47-48)

--- Direct juxtaposition of Head-Gen; syntagm marked purely

by L-to-H tone change on Head before low-tone Gen

--- Used in forming compounds (99-100); also genitive strategy for "intimate noun" Heads (kind of inalienable possession)

--- Pronominal-suffix possessives (intimate nouns) belong here

(ii) Analytical, with Prep kó "of": Head kó Gen (74)

--- Used for possession with non-intimate noun Heads

--- Patterning of Art (both types) in Gen construction (both types):

(i) Relational-affix genitive:

--- (a) Determinant: statement that compounds do not take Det at all (99); yet several exx. (49) do seem to show Det affix, occurring once only: Head-Gen-Det, e.g.

ears-hearing-Det ("ears which hear")

--- (b) Postclitic: stated to follow entire construction (51); thus numerous exx. passim: Head-Gen-postclitic

(ii) Analytical genitive (kó):

--- (a) Determinant: little data; one ex. with kó seems to have Det finally: "manner of preparation-Det" (49)

--- (b) Postclitic: little data; two clitic-final exx (53)

9) Nonconcord of V with full-NP Subj --- not applicable

--- Subj not coded on Verb

--- Subj pronoun omissible if clear in context (130); cf. p. 102 (usage of personal pronouns)

10) Verbal abstract: VN or Inf? --- probably neither

--- No real "infinitive" (explicit statement, 111); verbs do not code Subj

--- Verb always takes one of four suffixes: Imperfective, Perfective, Emphatic, Nominalizer (44-46)

--- Emph form repeats main verb for emphasis (46, 127); repeated form (Emph) seems nominal, but comes at end of VP, after Obj (125), hence questions of rection irrelevant

--- Nlzz (46) very productive, but limitations on its use; appears especially as Gen element in Head-Gen compounds to convey "Noun of (for) Verbing", yielding (apparently lexicalized) names for things; most exx. have no Obj, but two do show an Obj noun ("thing"), governed as in finite verb; will not count this as "verbal abstract"

--- Impf and Perf have both H-tone and L-tone variants of verb-base, with different uses (110-13):

H-Impf: independent clause
 L-Perf: independent clause and some subord uses
 L-Impf: as complement (no Subj) or in subord clause (with Subj)
 H-Perf: verb as Noun or Adj (no Subj), or after Aux (with Subj)

--- Summary: all Impf or Perf forms may take a Subj, hence not verbal abstract; Emph never has Obj; only Nlzz might count as Inf, but just 2 exx. given

11) Predicative particle? --- YES

--- Verb "BE" (114) has suppletive variants: o (sg), ya (pl)

--- Morpheme nÉ (76-77), called a "copula" but apparently a particle; cooccurs with verb "BE" in syntagm:

Subj BE nÉ Compl (77)

--- Homophonous (but not identical) to loc/instr Prep nÉ (76-77); discussion of historical link between three nÉ elements: "and", Prep, and copula (117 n. 18), but not homophonous "go" (115)

--- Other copular constructions, verbless: with and without nÉ, or with connective há (131-32); copular verb may also occur without nÉ (125, 117 n. 20)

12) Prepositional periphrastic (Prep + VN) --- NO

--- Verb nÉ "go" usable as future Auxiliary, homonymous to loc/instr Prep nÉ; but resemblance fortuitous (114-15, cf. 117 n. 18)

13) "DO" periphrastic ("DO" + VN) --- no mention, assume NO

--- No mention of dÉ- "do" among Aux verbs (114-15)

14) Adverbial clause = "and" + finite clause --- weakly (semantics)

--- Three clause-linking "and" words: nÉ, go, tÉ (63-66, 70); no Adv use shown for nÉ, go; but tÉ indicates either sequence or purpose (70)

--- A unique ex. (66, 132): "And (he) takes him and a child with-him", meaning "and there was a child with him" or "a child came along with him"; adverbial construal?

15) VN/Inf instead of finite main-clause form --- NO

--- No proper "verbal abstract" (see [10]); Emph and Nlzz forms not used in predications

--- Subj omissible in various contexts (130); but with or without Subj, identical verb forms are involved

16) Word-initial change --- no mention (34-35), assume NO

17) Extended use of kin terms ("Kin of Noun") --- a bit

--- Dictionary (190-229) indicates that bé-, the combining form of béem "child", also means "small"; 3 exx. in use "small" (191, 193, 219), 6 exx. as literal "child" (193, 205, 206, 207, 211, 214); two good exx.:

bé-duk = child-(of)-mortar "pestle" (195)

bé-tom = child-(of)-message "messenger" (222)

Georgian (Caucasus; Kartvelian = South Caucasian)
Aronson 1982; also Aronson 1972, Tschenkeli 1958, Vogt 1971

- 0) Basic categorial information
 - Noun: no gender, no article (1958:1); case and number coded suffixally (1958:2); case system is split-ergative (on tense) (1982:110-11, 136); slot sequence: Noun-Plural-Case
 - Verb (1971:80-87, 1958:347ff., 1982:41, 169-77) codes Subj and one Obj (direct or indirect); since 3-pers DObj marker is zero, bipersonal form can implicitly be tripersonal (1971:86); 1/2-pers Subj and Obj markers (also special 3-pers IObj) are prefixes; plural marker is suffix -t, used (ambiguously) for 1pl-Subj, 2pl-Subj/Obj; 3-pers Subj expressed by suffixes
- 1) Conjugated adpositions --- NO
 - Postpositions (1982:90-92) added to case-inflected personal pronouns just as to full Nouns (1982:244-45, 1958:145-46); explicit exx. (1971:76, 78) of indep Pron as Obj-of-Postp
- 2) Word order
 - SOV, but with considerable freedom (1982:47, 1971:221)
 - Postpositions (1982:90-92)
 - Adj-N, reverse rare (1982:68, 1971:220)
 - Gen-N, reverse rare (1982:68, 1971:220)
 - N-RC1 (1972:140, 1982:179-80 exx.)
- 3) Relative clause linker --- Relative Pronoun
 - Case-inflecting RelPron romeli-c of Indo-European type, formed from interrogative + -c (1982:179); RelPron codes function of HeadN in RC1
- 4) Relativization strategy/ies
 - No significant difference from IE construction, as in e.g. Latin or German (explicit statement, 1972:136); in Postp RC1, Postp is fronted with RelPron (1982:180), sandwiched between romeli- and -c; exx. of genitival RC1 (1958:201)
 - Standard syntagm is: HeadN RelPron RC1, in that order, though if HeadN immediately precedes clause-final verb in matrix clause, that verb may intervene between HeadN and RelPron+RC1 (1972:140)
 - In spoken language, the inflecting RelPron romeli-c is replaced by invariant ro, rom, which usually does not come clause-initially (1972:139, 141); but little data provided
 - Sometimes romeli can function as a Relative Adj modifying HeadN ("which man..."), yielding correlative-style RC1 (1972:140-41, with exx.); if so, element -c occurs suffixed to [romeli + Noun]
- 5) Special relative form of verb --- NO
- 6) Polypersonal verb --- 2 actants coded explicitly
 - Verb can be implicitly tripersonal, with zero-marked 3-pers actant
- 7) Infixing/suffixing alternation --- NO
 - If verb has Preverb, the Subj/Obj prefixes come between Preverb

- and Verb (1982:40-41, 463ff.); but no real alternation involved
- 8) Definite article in genitive embeddings --- not applicable
 --- No article; syntagm generally: Dept(gen) HeadN (1982:68)
 --- Rare possibility of N-Gen order (1958:56ff.)
- 9) Nonconcord of V with full-NP Subj
 --- Inanimate plural Subj usually has 3-sg verb, Animate plural usually 3-pl (1982:89)
- 10) Verbal abstract: VN or Inf? --- VN
 --- Verbal noun (or masdar) takes Obj in genitive (1958:71ff., 1982:47-48, 69, 135-36); Obj usually precedes the VN, just as with normal genitive (1958:72); construction explicitly compared to Arabic verbal noun (1958:72)
 --- Participles, too, take their Obj in genitive (1958:533)
- 11) Predicative particle? --- NO
 --- Syntagm: Subj Pred Cop (1982:66), with Pred in nominative; in 3-sg, Cop may optionally be special enclitic suffix -a (cf. 1958:43 for AdjPred)
- 12) Postpositional periphrastic (VN + Postp) --- no mention, assume NO
 --- VN may occur with Postp (1958:73), but no mention of use with Copula
 --- Participles enter into compound tenses (1958:553); but Ptcpl is basically an adjectival element
- 13) "DO" periphrastic (VN + "DO") --- no mention, assume NO
- 14) Adverbial clause = "and" + finite clause --- no mention, assume NO
 --- General-purpose "and" word is da (1982:49, 1971:212); no AdvCl use mentioned
- 15) VN/Inf instead of finite main-clause form --- no mention, assume NO
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- a bit
 --- Dictionary: Tschenkeli 1960-1974
 --- "Mother" (deda) used in compounds in sense "main" (e.g. "mother idea" = main idea); not the right type
 --- Several Gen-N compounds in švili "child": "death's child" = mortal man; "work's child" = worker; "sin's child" = poor fellow; also "knife's child" = small knife (wrong type)

Gilyak/Nivkh (Eastern Siberia; isolate [Paleosiberian])
 Jakobson 1971, Comrie 1981, Panfilov 1962(I)-1965(II), Kreinovich 1934
 [Thanks to David Peterson for indispensable help with Russian sources]

0) Basic categorial information

--- Noun: several cases (mostly spatial; paradigm Panfilov I:122), with "Absolutive" case covering Subj, DObj, IObj, and Possessor (Comrie 268); no gender (but numeral classifiers, 269); plural suffix -ku (used on N and/or finite V, but not consistently) (270); no article
 --- Verb: codes number, not person; only Imperative codes person/number (270); gerunds make two-way distinction of "2,3sg" (in r-) vs. "1sg and all plurals" (in t-)

1) Conjugated adpositions --- problematic

--- Construction "Obj - Postpos" is same as construction "Possessor - N": both have dependent in Absolute case (qua exponent of genitive) (Panfilov I:147)

--- There is one single set of personal pronouns, which can be used ad-verbally (e.g. as Object), ad-nominally (as Possessor or Obj-of-Postp), and autonomously (Jakobson 86); because can appear autonomously, perhaps ought not to count as clitics; yet some can reduce phonologically, conditioned by initial consonant(s) of next word (Jakobson 88); Jakobson speaks of the "prefix-root" bond as tighter than "root-suffix" bond, also says: "The prepositive pronoun ... [is] more firmly and closely joined to the following stem and thus becomes a kind of prefix" (87-88)

2) Word order (Comrie 269)

--- SOV (pretty rigid)
 --- Postpositions
 --- Attribute-Head order, including Gen-N and RCl-N
 --- Category "Adj" does not exist; stative verbs instead (Panfilov II:11)

3) Relative clause linker --- ZERO

4) Relativization strategy/ies --- PARTICIPLE, GAPPING

--- "Participle" modifies nouns, apparently its sole function (Panfilov II:134); participle lacks all nominal categories, e.g. does not inflect for case or number (though indicates verb plurality via reduplication, 134); apparently consists of verb stem (including such verbal categories as voice, tense, aspect, modality, 136)

--- Subject vs. Object RCl's differentiated only by mutation behavior (see [16] below) of RCl verb/participle; Subj RCl (for transitive verb) will have verb mutation (since verb is preceded by an object), Obj RCl will not (II:63-64)

--- No mention of oblique RCl's

5) Special relative form of verb --- PARTICIPLE

6) Polypersonal verb --- one actant (Obj), clitic (?)

--- Basically one actant (Obj), coded as clitic (?) or indep pronoun
 --- Normally no affixal person marking; plural marker -ku, same with

noun and finite verb (Comrie 270); only Imperative distinguishes 2sg, 2pl, 1pl(incl.)

--- Person/number marked (preverbally) with nonaffixal pronouns; in certain functions (Noun possessor or Verb object but not Verb subject), pronoun is recast as "prepositive pronoun" and can undergo phonological reduction (see [1] above) (Jakobson 87-88, Kreinovich 206); this would seem to justify clitic status for pronoun

--- Several verb forms functioning as cross-clause linkage markers ("Gerunds", also coordinative -ta/-ra) occur in two alloforms, beginning in r- and t-: the r-form codes 2sg/3sg Subj, t-form codes 1sg and all plurals (270); this is germ of Subj coding

7) Infixing/prefixing alternation --- no mention, assume NO

8) Definite article in genitive embeddings --- not applicable

--- No explicit statement re nonexistence of articles, but texts (Comrie, Jakobson) show none and grammars mention none

--- Genitive strategy: Possessor-Head, with Possessor in Abs case and Head showing mutation (see [16] below)

9) Nonconcord of V with full-NP Subj --- sort of

--- Plurality concord not consistent; may be marked on noun or verb or neither (or both) (Panfilov I:92, exx. 99)

10) Verbal abstract: VN or Inf? --- CAN'T TELL (?)

--- Various nominalization formations: Infinitive in -tox (Panfilov II:151); gerunds (but see [15]); also the finite-verb suffix -d' apparently can also serve to nominalize a verb (Panfilov I:116); Panfilov asserts infinitive and d'-nominalization retain same verbal rection as finite verb (II:152, 153)

--- BUT, how could one distinguish VN-rection from INF-rection? Genitive and accusative both are case-marked identically (Absolute), and both involve identical mutation of the following head (see [16]); Gen-N and Obj-V syntagms thus formally identical, hence the notions "VN" and "Infinitive" would seem indistinguishable [? my own argument]

11) Predicative particle? --- sort of

--- Syntagm is: Subj Pred-ra ("Subj is Pred"), with Pred in Absolute case; ra is a predicative particle (Jakobson 96, Panfilov I:130) usable with (apparently) any predicate, nominal or verbal; shows -ra/-ta alternation (2/3sg vs. other)

--- Jakobson has one example without particle (80)

12) Postpositional periphrastic (VN + Postp) --- no mention, assume NO

13) "DO" periphrastic (VN + "DO") --- no mention, assume NO

14) Adverbial clause = "and" + finite clause --- not applicable (?)

--- Gerunds used almost exclusively instead of finite subordinate clauses (Comrie 270)

--- There is a "coordinative" suffix [my term] -ta/-ra (distinct from homophonous predicative particle in [11] above), such that Verb-ta functions rather like a gerund but appears in independent clauses; Comrie glosses this as "AND" (270), but Panfilov (II:116-17) calls it a

"repetitive" (transl.?) suffix; if Verb-ta were truly parallel to gerund (i.e. used to chain clauses), then gloss "and" would be overtranslating; would seem not to be a real conjunction (contrast true noun-conjunction -xe "and", Panfilov I:165ff.)

15) VN/Inf instead of finite main-clause form --- weakly

--- Heavy use of gerunds of several types, apparently both for notional subordination and to carry forward the narrative, though semantic distinctions not clear (Comrie 270); texts (Comrie, Jakobson) show gerunds being used narratively (cf. Jakobson's comments to text, 99-101); but these apparently do not fulfill any prototypical functions of "VN/Inf"

16) Word-initial change --- YES, multicategorical

--- With very few exceptions, transitive verbs are underlyingly fricative-initial (Jakobson 86), and nouns are underlyingly stop-initial (93); these are then subject to initial mutation

--- Basic mutation pattern (simplified and modified from Comrie 267): in appropriate syntactic contexts (below), word-initial segment becomes

(i) Fricative after word-final stop or vowel or /j/ (affects Noun)

(ii) Stop after word-final fricative or nasal or /l/ (affects Verb)

--- Syntactic contexts triggering mutation: Head regularly undergoes initial mutation (so as to conform to basic mutation pattern, above) in the following two Dept-Head syntagms:

Dept-Head = Obj-V, Gen-N (Comrie 268)

--- Fricative-initial transitive verbs stand opposed to stop-initial intransitive counterpart

17) Extended use of kin terms ("Kin of Noun") --- apparently not

--- Checked in Nivkhsko-Russkij Slovar'; nothing evident

Classical Greek (ancient Greece; Indo-European)

Smyth 1956; also Dover 1960

[My thanks to Andrew Riggsby for answers to several queries]

0) Basic categorial information

--- Noun: marked for gender (m/f/n), number (sg/du/pl), case, conflated in a series of complex suffixal paradigms (declensions) (44ff.)

--- Verb: marks person/number of Subj only, coded suffixally (107ff.); numerous distinct sets of personal endings depending on tense, mood, voice

--- Article: proclitic definite article, coded for gender/number/case to agree with noun (94, 286ff.); no indefinite article

1) Conjugated adpositions --- NO

--- Preps govern the oblique cases of pronouns (366); some Preps are themselves proclitic (41)

2) Word order

--- SV clearly favored over VS; OV somewhat commoner than VO (Dover 25, 30)
 --- Prepositions almost exclusively (Smyth 368-69)
 --- Adj-N more common than N-Adj (Smyth 293)
 --- Gen-N, N-Gen both very common (Smyth 294)
 --- N-RC1 (clear though not explicitly stated; AR, p.c.); also correlative (Smyth 560-61, 570)

3) Relative clause linker --- Relative pronoun
 --- Inflects for case/number/gender (97); initial in clause (560)

4) Relativization strategy/ies --- GAPPING
 --- Normal RC1 type is externally headed: NP gaps, shows up as RelPron clause-initially
 --- In Prep RC1, Prep fronts along with RelPron [AR, p.c.]
 --- In Gen RC1, Gen RelPron is fronted; possessee NP survives, usually with article, and usually fronted [AR, p.c.]; thus:

the men [WH(Gen) the-house(Nom) has been destroyed]
 "the men whose house has been destroyed"

--- There is a fairly "pure" correlative type, involving two coordinate clauses; also a different type described as follows: "The antecedent [is] taken up into the relative clause ..., the relative agreeing adjectively with its antecedent" (570); RelPron still initial in RC1, but antecedent need not be adjacent to it and need not explicitly be referred to in both clauses (570); this would appear to be Internally headed; ex. (HeadN = "virtue"):

if is [WH(Acc) you formerly spoke-of virtue(Acc)] real
 "if the virtue which you were speaking of before is real"

5) Special relative form of verb --- NONE

--- Finite RC1s use ordinary verb forms; participles may be used adnominally (455), but also adverbially (456-57)

6) Polypersonal verb --- 1 actant coded (Subj) suffixally

7) Infixing/suffixing alternation --- not applicable

8) Definite article in genitive embeddings

--- Each NP retains its own article; various orderings possible (294):

(a)	[the [the Poss] Head]	(very common)
(b)	[the Head] the [the Poss]	(less common)
(c)	[the Poss] [the Head]	(emphasize possessor)
(d)	[the Head] [the Poss]	(very common)

--- Note especially the "sandwich" construction (a); only restriction is that the two adjacent articles should not have the same form (294)

9) Nonconcord of V with full-NP Subj --- a little

--- Basic rule is full concord (262); neuter plural Subj construed as

collective, hence normally takes singular verb (264)

--- Pindaric construction: with certain copular verbs, occasionally find masc- or fem-plural Subj with a singular verb; order here is usually Verb-Subj, so that "the subject is still undetermined" when verb is uttered, the plural Subj coming as a kind of "afterthought" (264); rare, poetic (?)

10) Verbal abstract: VN or Inf? --- INFINITIVE

--- Infinitive is the only productive deverbal substantive (107, 437); it takes its Obj in same case as finite verb, and never in "objective genitive" (438); Inf does not itself inflect for case, but may be accompanied by neuter Article which does show case inflection (438)

11) Predicative particle? --- NO

--- Predicate nominal (Adj or Noun) takes same case as Subj, hence Nominative in independent clauses (257, 261); copula omissible (261); no particle

12) Prepositional periphrastic (Prep + VN) --- no mention, assume NO

--- Inf (with Article) may be governed by Prep (438, exx. 451-52), but no indication of our construction; periphrastic tenses discussed (436-37), again no mention of construction

13) "DO" periphrastic ("DO" + VN) --- no mention, assume NO

--- Discussion of Inf as Obj of finite verbs (443-53 passim), but no mention of Inf as Obj of "DO"; no periphrastic tense with "DO" (436-37)

14) Adverbial clause = "and" + finite clause --- apparently YES

--- There are two clause-level "and" conjunctions (kaí, -te); discussion of Parataxis ("coordination in place of subordination") indicates the phenomenon is common in Attic prose, inter alia with kaí (485-87)

15) VN/Inf instead of finite main-clause form --- a bit

--- Only predication use of Inf seems to be to express commands or wishes (448)

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- NO

--- Liddell & Scott dictionary mentions constructions with "son-of" as being Hebraisms, found specifically in the Septuagint; hardly any relevant examples in dictionary: "wine" = "child of vine"; poetic reference to "the mountain-rock's child, Echo"

Hausa (West Africa; Chadic [Afroasiatic])
 Kraft and Kirk-Greene ("KK") 1973; also Kraft 1963,
 Cowan and Schuh ("CS") 1976, Howeidy 1953, Abraham 1962

0) Basic categorial information

--- Noun: takes possessor suffixes (46-47) following linker -n/r (n for masc-sg noun and all plurals, r for fem-sg nouns; longer forms are na, ta):

Noun-n/r-Sfx ;

masc/fem gender, partly covert, neutralized to masc in plural (27, 42); complex plural marking involving tone change, suffix, and/or change to stem-final vowel (121ff., 242ff.)

--- Verb: preceded by Person/Aspect subject pronoun (37), here analyzed as Subj proclitic, normally obligatory (but see [6] below); optionally followed by one Obj pronoun (direct or indirect) (74-75), here analyzed as enclitic; with two objects (noun or pronoun) order is always V-IObj-DObj, with DObj pronoun now appearing in distinct Independent form (75)

--- Article: linker -n/r can indicate definiteness when suffixed to noun with no following noun (52-54, cf. CS 101); sometimes the postnominal word Dîn is used instead; but a bare N can itself be definite with no explicit mark

1) Conjugated adpositions

--- "Prepositions" not presented as such, but as "relaters" (85) and "relational nouns" (87-89), chart p. 90; not many

--- Of "relaters", gà/gàrē "to, for" takes DObj pronoun (an enclitic form) (per exx. 86), whereas dà "with" takes Indep pronoun (nonenclitic form) (95, 150)

--- Relater à "at, in, on" almost always precedes some positional word, notably relational nouns (CS 58); thus precludes pronominal Obj

--- IObj formed by wà + Noun, or by fused pronominal forms ma+Pron (74), but some dialects use mà instead of wà; plausibly all are allomorphs of a "Dative Prep" (my own analysis); if so, it's a conjugated Prep, using Pron-possessor endings (74, 47)

--- Some (all?) "bare" relational nouns used in adverbial function with no NP Object, e.g. "inside" (CS 58)

--- Here will analyze the "relaters" as prepositions

2) Word order

--- SVO (no direct statement, but see exx. passim and CS 196)

--- Prepositions (85)

--- Adj-N (129); sources disagree about existence of a class "Adj": KK (129) views Adj as subclass of true nominals, CS (312) takes it as its own class; will follow CS; syntagm is:

Adj-n/r N (or, appositionally): N Adj (without n/r)

--- N-Gen (41)

--- N-RC1 (106-7)

3) Relative clause linker --- INVARIANT

--- Either dà or wandà (106); wandà is gender-inflected per head; dà (invariant) must be preceded by linker -n/r on HeadN; hence take wandà as the demonstrative wa- (51) functioning as appositional restatement of

head (my own analysis):

Head-n dà RC1 (for masc head)
= Head wa -n -dà RC1

--- dà has many uses (188): "with/and", "when", RC1, to express "have" ("BE dà Possessee"), DObj

4) Relativization strategy/ies --- basically GAPPING

--- Sometimes verb assumes special RC1 form (see [5]), marked prefixally (CS 104)

--- Technique appears to be Gapping, per exx. (CS 107); exx. given only of Subj, DObj, and Time/Place; no discussion of Possessor RC1, but one ex. in Kraft 1963 (II:91), showing Copying:

son male REL fineness-his not it has limits
"a son whose fineness has no limits"

--- Status of Prep RC1s and IObj RC1s unclear; 3 exx. indicate different things:

(a) Relativizing on "have" (= "be with") requires resumptive pronoun: "The cattle that I am with them" = "The cattle that I have" (CS 215)

(b) Clefting (using Rel form of verb) on an IObj strands the "indirect-Obj Prep" wà in situ and gaps its Obj (CS 196):

Shēhù nē na kē kāwō wà ____ kāya-n gōnā
PN Ptcl I Cont.Rel bring IObj ____ goods-n farm
"It's Shehu I'm bringing the farm goods to"

(c) Ex. with relational noun shows apparent gapping of Obj of "Prep" (KK 127, 307 (from exercises)):

inā gidājēn dà mukà bar kāya-m- mù à ciki jiyà
where compounds REL we.REL leave load-LK-our inside yesterday
"Where are the compounds in which we left our loads yesterday?"

5) Special relative form of verb --- YES, sometimes

--- Only for Continuative and Completive aspects (not Habitual and Future), only in positive; expressed by special Pers-Asp prefixes (104-5, table of prefixes 226)

--- These special forms also used for clefting and WH-questions (108-9)

6) Polypersonal verb --- 2 actants (Subj, Obj), coded as clitics

--- Subj proclitic, a single Obj enclitic (see [0] above); but (NB) in Continuative aspect, Pers-Asp prefix may be dropped if full-NP subject (CS 197)

7) Infixing/suffixing alternation --- no mention, assume NO

8) Definite article in genitive embeddings

--- Two genitive syntagms (41):

/

- (a) Head-n/r Dept: HeadN and Dept must be adjacent
- (b) Head na/ta Dept: something separates HeadN and Dept (42; CS 243)

--- In (a), since genitive linker is same morph as "article" (see [0]), no "article" qua article can appear on Head (my own argument); thus only Dept can be marked "definite"

--- In (b), head apparently can be marked Def, per ex. (Howeidy 28):
 sarki-n na Kano "the Emir of Kano" (sarki "chief")

9) Nonconcord of V with full-NP Subj --- no mention, assume NO

10) Verbal abstract: VN or Inf? --- VN (and apparently INF)

--- Grammars present a "verbal noun"; two types, one ending in -wā, one with no ending (CS 179-81, KK 98-101)

--- Type in -wā precludes DObj (CS 182); if DObj appears, verb reverts to regular (non-VN) form

--- Endings type takes Obj in genitive (CS 182); but if IObj appears (order: V-IObj-DObj), verb reverts to regular (non-VN) form (KK 101)

--- This "reverted regular form" would appear to govern Obj like finite verb, hence Inf; but no explicit statement

--- Usable like any Noun, ex. "Traveling is better than sitting" (CS 179); appears as complement of Aux (KK 158-59); appears in Continuative Aspect verb forms after Aux, cf. [12] (CS 181-82)

--- In prototypical EQUI contexts, use Subjunctive (KK 63, 167-68)

11) Predicative particle? --- NO (depends on analysis)

--- Three constructions:

(a) Subj PredN/PredAdj nē/cē (cē for fem-sg, nē otherwise)
 (KK 32-33, ex. 132, cf. CS 313)

(b) Subj Aux WITH N-of-quality (CS 69-70: "The tree has bigness")

(c) Subj DO N-of-quality (CS 166: "The girl does beauty")

--- nē/cē not a verb (CS 48); but fact of gender/number inflection makes it un-particle-like; analyze here as pseudo-copula [OG: cf. similar constructions in Semitic, Egyptian]

12) Prepositional periphrastic (Prep + VN) --- in a way

--- Inflected Aux element -nā (3sg yanā, etc.) used thus (CS 181):

yanā + with

(a) Locational: yanā + LocAdv

(b) Have: yanā + with + N

(c) Continuative Aspect: yanā + VN

--- Can have non-deverbal "action noun" instead of true VN

--- Thus effectively have: locative Cop + VN (see KK 93-95, CS 181-82)

13) "DO" periphrastic ("DO" + VN) --- NO

--- "DO" not mentioned in list of Aux verbs (CS 242-43); have a construction (CS 166) exemplified by "do beauty" (= be beautiful), but no VN involved

- 14) Adverbial clause = "and" + finite clause --- not applicable
 --- No clause-level "and"; clauses conjoined asyndetically (Howeidy 191), no discussion in KK or CS; AdvCls expressed by various conjunctions (KK 87-89, 185-88)
- 15) VN/Inf instead of finite main-clause form --- no mention, assume NO
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- YES
 --- "Son"/"daughter" (Dā) is discussed in grammar, used for place of origin, occupation, stylized expressions (193-94); e.g. "son of stick" = policeman; "son of table" = petty market trader
 --- Further exx. in Abraham 1962 (large dictionary), esp. s.v. "son", "mother": "mother of gown" = body of gown, "mother of money" = capital, "son of beans" = a bean dish; numerous others passim

Hawaiian (Pacific; Polynesian [Oceanic = Eastern Austronesian])
 Elbert & Pukui 1979 (also Hawkins 1979, 1982)

0) Basic categorial information

- Noun: no inflectional cases, though Subj and Obj expressed via case-marking Preps (131ff.); no gender; plural marked on a few nouns by vowel-lengthening (106), otherwise expressed by plural article (156) or separate plural word (162-63)
 --- Verb: no inflection for person/number; can mark plural by reduplication (66) or prefix (75)
 --- Article: definite (ka/ke) and indefinite (he) article (154-58, 1982:1ff.)

1) Conjugated adpositions --- NO

- Preps apparently take the single standard series of (multi-purpose) pronouns (107); the Acc/Dat/Oblique Prep ia is written together with singular pronouns (38, 133), but little or no phonological change seems meant

2) Word order

- VSO (39-40)
 --- Prepositions (131ff.)
 --- No category "Adj"; use stative verbs (43-44, 49)
 --- N-Gen, also Gen-N (136ff., 143)
 --- N-RC1 (1982:108; Hawkins calls RC1s "subordinate clauses")

3) Relative clause linker --- ZERO

4) Relativization strategy/ies (1982:90ff., 108ff.) --- usually COPYING

- All RC1s: certain preverbal tense/aspect markers change to medial form (Perfective/Inceptive ua ==> i; Present ke ==> e) (58-60)
 --- Subject RC1: delete subject NP (no ai appears) (1982:90)

--- Non-subject RC1 (including Obj RC1): replace embedded relative NP (and its Preposition!) with anaphoric particle ai, which does not occur "in place" but is positioned just after Verb (41); but ai will not appear if certain postverbal adverbial markers occur (ana, ala, nei) (1982:90ff.)

--- Optional (but preferred) in non-subject RC1: Subj of RC1 may be raised into matrix clause to become Possessor of Head-Noun (1982:112):

ua kōkua ka māka'i i ke keiki
Perf help the policeman Obj the child
"The policeman helped the child"

==> ua ho'i mai ke keiki [i kōkua ai ka māka'i]
Perf return the child Perf help Anaph the policeman

OR: ua ho'i mai ke keiki a ka māka'i [i kōkua ai]
of the policeman Perf help Anaph
(lit. "The child of the policeman [that (he) helped him] ...")

BOTH = "The child [that the policeman helped] returned"

--- NB: this "genitive subject" ("policeman's") belongs syntactically to matrix clause, not to RC1: like any Gen, it can be preposed (Gen-N) to its Head (exx. 1982:113-14), and as such is no longer even adjacent to the RC1, as if structure (NB: preserving same meaning!) were

"The policeman's child [that (he) helped]"

--- No discussion of possessor relatives

5) Special relative form of verb --- sometimes

--- Certain tense/aspect markers change (see [4])

6) Polypersonal verb --- ZERO actants coded

--- Only a single set of pronouns (107)

--- Subj and Obj pronouns typically marked by Prep Phrases (131-32); find exx. (108) of Subj pronoun with no Subj-Prep, but no indication that these pronouns are clitics; pronouns often omissible anyway (108)

7) Infixing/suffixing alternation --- not applicable

8) Definite article in genitive embeddings

--- Normally N-Gen, but preposed order (Gen-N) also possible

--- Genitive preposition a/o is used; Article appears on both Nouns (139ff.; no discussion, but clear from examples); thus:

(i) Art N [a/o Art Gen]
OR: (ii) Art [a/o Art Gen] N (Gen preposed, 143-44)

--- In (ii), initial Art fuses with a/o marker

9) Nonconcord of V with full-NP Subj --- not applicable

10) Verbal abstract: VN or Inf? --- INFINITIVE

--- "Nominalization" (79-80) involves adding particle 'ana to verb,

with resultant gerund-like nominal generally preceded by Article, Dem, or Poss (1982:63); one ex. (64-65) of EQUI type ("forget to V")

--- Subj of nominalization occurs in Genitive (1982:63), Obj in usual form (exx. passim); hence counts as Infinitive

--- Nominalization (per exx.) doesn't appear to take preverbal tense/aspect/mood markers, hence seems distinct from finite clauses; but no explicit statement

--- 'ana sometimes omissible (64)

--- Verbs can also take finite clauses as complements (Hawkins 1979:28-29); specifically, the preverbal mood marker e (Elbert: "intensive", 58, 61) occurs "whenever one verb is the complement of another" (1982:44; Hawkins calls this the "infinitive" use)

11) Predicative particle? --- sort of (for definite predicates)

--- Order is: PredN Subj, with no copula expressed (1982:70); if PredN is definite (takes Def Article), its Article is preceded by 'o' (1982:55)

--- This 'o' is also used (before pronouns [esp. 3-sg] and proper nouns) to mark Subject NP, and as such is termed a nominative preposition (131-33, 1982:55); also used with sentence-initial topics (1982:55, Hawkins 1979:58-59), and with appositive NPs (1982:55)

12) Prepositional periphrastic (Prep + VN) --- no mention, assume NO

13) "DO" periphrastic ("DO" + VN) --- no mention, assume NO

14) Adverbial clause = "and" + finite clause --- apparently YES

--- Conjunction a glossed "and, until, like" (164); ex. (165):

"Work and [i.e., until] the work is finished"

--- Dictionary (Pukui & Elbert 1971) translates it as "when, at the time when, until, to, as far as, and, or (rare), and then, but"

15) VN/Inf instead of finite main-clause form --- no mention, assume NO

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- NO

--- No good examples found in dictionary (Pukui & Elbert 1971); a few zoonyms; also "child-go" = traveler

Hittite (ancient Anatolia; Anatolian [Indo-European])

Friedrich 1974; also Luraghi 1990

[My thanks to Gary Holland for clarification on several points]

0) Basic categorial information

--- Noun: gender (42-43: common (m/f) vs. neuter) usually a covert category, reflected in concord phenomena; number and case coded suffixally, conflated in several declensional paradigms (43ff.); in Old Hittite, nouns may take Possessive-Pronoun suffixes, with both noun and

suffix taking case inflection, the suffix in same case as noun (65, 133-34); no articles

--- Verb: suffixal conjugation for Subj (77-78), conflating pers/numb/tense; also have a clause-initial block of enclitic particles including Subj, DObj, and IObj markers (maximum two), subject to various cooccurrence restraints (147-48); these clitics hosted by the verb only fortuitously, when it occurs clause-initially

1) Conjugated adpositions --- YES

--- Five locational postpositions (Old Hittite) may take Poss-Pron suffixes as Pronominal Obj (133-34); pattern identical to that of possessed nouns

--- For three such, Postp acts as Nom/Acc-Neuter noun, with suffix following suit; for other two, Postp appears in Dat/Loc-sg form, with suffix also Dat/Loc

--- Best conceptualized as e.g. [underness-its]-Dat/Loc = "under it"

2) Word order

--- SOV (1990:36)

--- Postpositions (1974:129)

--- Adj-N (1990:139 n. 4)

--- Gen-N (1974:122); if HeadN is ideogram, may find N-Gen (artificial?)

--- RC1: correlative; normal order is RC1 MainCl

3) Relative clause linker --- not applicable (correlative)

4) Relativization strategy/ies (167-69) --- Correlative

--- Two independent, coordinate clauses, with a shared participant in both; usually RC1 precedes MainCl

--- In first clause ("RC1"), the conceptual HeadN appears as such, normally in its usual place within the clause (i.e. not fronted), singled out as HeadN by its accompanying Relative Adjective kuiš; RelAdj fully inflected (68-69), agrees with HeadN, pre-specifies HeadN as being topic of (following) MainCl; may either precede or follow HeadN, depending on conceptual (in)definiteness of HeadN

--- In second clause ("Main clause"), HeadN is replaced by ordinary anaphoric pronoun, pointing back to participant identified as "HeadN" in RC1

--- Syntagm: you [WH man] saw, I his father know
Paraphrasable as: Which man you saw, I know his father

5) Special relative form of verb --- NO

6) Polypersonal verb --- only Subj coded

--- Obj enclitics occur in initial particle chain, but only fortuitously will Verb (normally clause-final) serve as host-word for clitics

7) Infixing/suffixing alternation --- not applicable

8) Definite article in genitive embeddings --- not applicable

--- No article

--- Basic syntagm: Dept-gen Head (122);

also syntagm: Dept-gen Head-poss ("des Mannes sein Kopf")

- 9) Nonconcord of V with full-NP Subj --- not really
 --- Basic rule is full concord; Neuter-plural Subj takes verb in sg (118)
- 10) Verbal abstract: VN or Inf? --- both; Inf dominant [GH, p.c.]
 --- "Infinitives" (two morphologically conditioned allomorphs, 112-13), presented as "richtige Infinitive unserer Art" (142); i.e. accusative rection of Obj (1990:132 n. 49), though Obj often raised to matrix clause (143); Inf may function as purpose adverbial, noun-clause substitute (1990:132 n. 49), complement of Adj (1974:143), i.e. normal infinitival functions
 --- "Verbal noun" (again, two allomorphs, 112, 142), construed nominally, i.e. with genitive rection of Obj (ex. 142)
- 11) Predicative particle? --- NO
 --- Syntagm is: Subj PredNom/Adj (Copula);
 Cop usually zero in present tense, explicitly present in past tense (117-18)
- 12) Postpositional periphrastic (VN + Postp) --- no mention, assume NO
 --- Other periphrastic tenses (111, 136-38), but not with Postp
- 13) "DO" periphrastic ("DO" + VN) --- no mention, assume NO
 --- Verb "do" is iya- (92)
- 14) Adverbial clause = "and" + finite clause --- NO
 --- Several clause-linking "and" words (154ff.): -(y)a has no adverbial use; Classical Hittite clause-initial nu basically a main-clause conjunction, only non-main-clause use is in "because" clauses (159), but nu is not itself the word "because"; enclitic -ma (a weak "but") may occur in protasis of "if" clauses (162), but is not itself the word "if"
 --- A variety of Adv-clause conjunctions (163ff.)
- 15) VN/Inf instead of finite main-clause form --- NO
 --- Action carried forward in narrative with chained nu-clauses, rendered as "and-then" (155-56); but always fully finite verbs
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- apparently not
 --- No examples apparent in Friedrich 1952 (dictionary), outside of Akkadogram DUMU šipri "son of sending" = "messenger" = Hittite haluqatalla- (monolexemic) (269)

Hixkaryana (northern Brazil; Carib)
Derbyshire 1979 (cf. Derbyshire 1985)

0) Basic categorial information

- Noun: no cases, gender, or article (126-27); plurality coded by "collective" marker (126; optional for humans, forbidden otherwise), usually independent particle (84) but sometimes a suffix (100); Possessed nouns (44, 83) take possessive prefix and a "possessedness" suffix which additionally conflates number and tense-of-possession (96-100)
- Verb: Subj and Obj person coded prefixally (conflated) (145-46); "collective" number (of Subj and/or Obj) coded suffixally, once only, with suffix conflating collective/tense/aspect (136, 145)
- NB: almost no adnominal modification (44); rather, heavy use of paratactic constructions (26, 33, 39, 46), e.g.

Concerning-it I-wrote-it to-him, to-there his-going Denomlizr,
 Waraka his-going Denomlizr, Manaus to his-going Denomlizr
 "I wrote to him about it, so that he, Waraka, would go there,
 to Manaus" (33)
 ["Denominalizer" makes purpose clause by adverbializing
 a nominalized verb-form (29)]

1) Conjugated adpositions --- YES

- Postp takes "possessive" prefixes or full-NP but not both (43, cf. 96-97 for prefix paradigm); differs thereby from Gen-N (see [8] below)

2) Word order

- OVS (40), with flexibility re fronting; Obj almost always directly precedes V (87)
- Postpositions (43, 81-82, 154)
- "Adjectives and adverbs form one class" (153), called "Adverb" (41), but not used adnominally: Adnominal modification is essentially restricted to numerals (44); hence no adnominal Adj or adnominal RC1 (see [4] below for rough equivalents)
- Gen-N (45, 96)

3) Relative clause linker --- not applicable

4) Relativization equivalency strategy/ies (26)

- A) Various nominalizer suffixes (see [10] below) create equivalent of headless RC1s (91ff., 165ff.): "he who VERBs", "that which is VERBed", "thing/time/place associated with action", etc.; the construction is inherently headless
 - If used appositionally to another NP (with intonational break), may approximate to "adnominal RC1" (26), but neither NP would be "Head"; nothing said about ordering of NPs (one ex. shows order "Noun Nominalized-verb", 26)
 - Adj/Adv can be nominalized with suffix -no (169, cf. 167), e.g. "one that is good"; used appositionally, might approximate to adnominal adjectival modification (no explicit statement)
- B) Another strategy: an independent equative sentence can serve semantically somewhat like RC1 (26, 36-37) (like correlative RC1?);

e.g.:

he-came visitor; adult-man Devalued that-one (he-was)
 "A visitor came; he was an old man"
 (= "A visitor came who was an old man")

this type can approximate to Adj-N "modification" (45)

--- But no clearcut "relative" construction per se

5) Special relative form of verb --- not applicable

6) Polypersonal verb --- 2 actants (Subj and Obj)

--- Prefixal coding of person; number subsumed in verb suffix (145-46)

7) Infixing/suffixing alternation --- no mention, assume NO

--- Subj/Obj conflated (146); Subj/Obj prefix apparently always initial slot in Verb

--- In negation, special "Neg-Adverb" form of V precedes Copula, with Obj recast as Possessor of NegAdv (48, 24-25):

Positive: deer I.it-ate "I ate the deer" (48)
 Negative: deer its-NEG.eating I-was "I didn't eat the deer";

this splits Obj from Subj and repositions Obj, but "BE" counts as a full verb

8) Definite article in genitive embeddings --- not applicable

--- No articles; genitive syntagm is: Possessor his-Head-ri;
 possessor prefix obligatory, formally similar (but not identical) to "Obj prefix" of transitive verb (i.e. Subj/Obj prefix when Subj is 3-pers); -ri expresses possessedness (and tense and number) (45, 96-100)

9) Nonconcord of V with full-NP Subj --- sometimes (number nonconcord)
 --- A "Collective" Subj or Obj NP may fail to have Coll-marked verb; nonconcord rare for Subj, commoner for Obj (145-46)

--- When Trans Verb (3pers --> 3pers) is immediately preceded by a full-NP (either Subj or Obj), the 3-->3 prefix (normally y- or n(±)-) assumes a zero allomorph if the verb stem is consonant-initial; if this preverbal NP is the Subj, the zero allomorph will occur only when the verb is preceded by pause (87-88, 147). The phenomenon is common for Obj, rare for Subj. A somewhat similar rule holds for intransitive verb (147). Given the heavy phonological conditioning, it is unclear whether this zero should represent "nonconcord", though (significantly?) the person-prefix paradigm has no other zero allomorphs (146-47).

10) Verbal abstract: VN or Inf? --- VN

--- "Derived nominals" (23ff., 91-94) formed by nominalizer suffixes (165-68); person, number, tense still coded, but all categories marked as with nominal possession: tense recast as tense-of-possession, intrans Subj and trans Obj recast as possessor, trans Subj recast as Postp Phrase with wya "to, by" (23, 91)

--- Also "pseudo-nominals", formed with adverbializer suffixes (176-78) and functioning adverbially (exx. include purpose-EQUI contexts, 29); argument NPs marked as with derived nominals

11) Predicative particle? --- YES

--- Usual syntagm: Complement me Copula Subj (35); Comp always followed by Postp me "denominalizer" (34), e.g.

my-brother me he-is this-one "This is my brother"

--- General function of me (102-3): convert N to Adv, typically in meaning "in the form of", "as", e.g.:

[club me] [wood occupied-with] I-was
"I used the piece of wood as a club";

also in purpose clauses (28-29) and manner clauses (28), e.g.:

[loudly my-singing me] I-take.a.bath
"I take a bath singing loudly"

--- Also have equative (verbless) construction, with (NB) no particle (36-37)

12) Postpositional periphrastic (VN + Postp) --- possibly

--- Normal construction for "want to V" involves Postp xe "desirous of" following derived nominal (DervN) of Verb (32, cf. 107):

[poss-DervN xe] COP "be desirous of (one's) VERBing"
= "want to Verb";

syntax is exactly right, semantics a bit marked

13) "DO" periphrastic (VN + "DO") --- in a sense

--- No mention of any DO-periphrastic of the normal type

--- Extremely rich use of ideophones (exx. 190-91), typically as DObj preceding ka "say, do" (80, 82); construction thus parallels that of direct speech (80, 4)

--- But: "Most actions that can be expressed by an inflected finite verb form can also be expressed by a non-inflected ideophone [!]" (190); ideophone might thus almost count as a kind of suppletive VN [? my interpretation]; taking ka as "do" would then yield desired syntagm

14) Adverbial clause = "and" + finite clause --- not applicable

--- No "and" word (45); use juxtaposition

15) VN/Inf instead of finite main-clause form --- no mention, assume NO

16) Word-initial change --- conceivably

--- Note that the 3-->3 Subj/Obj verb prefix has different (non-zero) allomorphs depending on whether a full-NP Obj precedes (87, 146); construable as syntactically controlled word-initial change??

17) Extended use of kin terms ("Kin of Noun") --- seemingly not

--- No examples readily apparent in "Lexicon" section (192-98)

Hottentot (Nama) (southern Africa [Namibia]; Khoisan)
Hagman 1977

0) Basic categorial information

--- Noun: inflects for person/gender/number using a conflated "pgn" suffix; can have non-3-person marking in e.g. "you people" (22, 41ff.), but pgn suffix can also occur as enclitic to other parts of speech; also case inflection, covering a single "subordinative" case -à (56), used at clause-level for non-subject NPs and for demoted subjects, also for Obj of certain postpositions (57); also traces of an "agentive case -i (56, 105); slot sequence is: N-pgn-Case

--- Verb: optionally inflects for Obj only (sometimes for two objects) (79-81); inflection is a modification of pgn suffix; for illusion of subject inflection see [6] below

--- Article: none stated as such; so-called "indefinite gender" can supersede and replace a noun's lexical gender, indicating "some X or other" (24); demonstrative //náá (basically adnominal) is often translated "the", is commoner than English "that", and seems to indicate a degree of deixis intermediate between English "that" and "the" (38); but will not reanalyze this as "hedged article"

1) Conjugated adpositions --- NO

--- Pronouns are independent words (44, paradigm); syntax of PostpPhrase is specified for full-NPs only (101-2); ex. 134 ("between us")

2) Word order

--- SOV basic (60, 75); but much reordering possible (107ff.; see [6] below)

--- Postpositions (101-2)

--- Adj-N (30)

--- Gen-N (36-37)

--- RCl-N (124)

3) Relative clause linker --- ZERO

4) Relativization strategy/ies --- GAPPING or COPYING

--- (NB: process is oddly presented in grammar book (124-25), but apparently is as follows:) Ordinary finite clause (minus its "declarative" particle ke, 121) adjoins to left of HeadN; or else HeadN is absent but its pgn suffix (now enclitic to RCl verb) survives

--- Subj RCl has gapping; gapping strongly preferred in DObj or IObj RCl (but other roles may gap too, and if so Postpos will gap along with its NP); resumptive pronoun copy is common with Postpositional or Possessor relatives (but not universal, apparently; little explicit information given)

--- A "pgn-headed" RCl may follow a noun; taken as apposition

5) Special relative form of verb --- not exactly

--- Deletion of declarative ke common to all subordinate clauses (121)

6) Polypersonal verb --- 2 Object actants maximum coded suffixally

--- Subj not marked on V; Obj-suffixes (modified pgn suffixes) may

appear only if full-NP object (or independent pronoun) absent; both DObj and IObj suffixes may cooccur (V-IObj-DObj) (79-81)

--- Whenever any element X is fronted, it normally displaces Subject, which must move or delete but leaves behind a copy of its pgn suffix encliticized to X; if X is the verb, this yields "accidental" cliticization of pgn suffix to verb, creating illusion of "subject-marked verb" (110-11)

7) Infixing/suffixing alternation --- apparently not (puzzling)

--- Obj-suffix (see [6]) is presented as part of "outer layer of derivational suffixes on the verb", in same fixed slot as Passive, Reflexive, Reciprocal (77ff.); i.e., implicitly as part of verb stem

--- Perfective aspect (66-67) involves placing Aux "BE" after verb stem; unclear if "BE" counts here as a separable verb or as part of a single verbal complex; also unclear where Obj should go, i.e. whether

(i) V-Obj BE or (ii) V.BE-Obj ;

no explicit discussion of positioning of Obj (and no relevant exx.);

note that (i) might perhaps be taken to involve "infixing"; there is one example given of the form "V-Rflx BE" (67), which is structurally parallel to (i) and might argue for (i)

--- There are also compound verb-roots formed by "Verb+Postp" (70-71), presumably functioning as a fused unit (i.e. no possibility of "infixing" Obj between Verb and Postp); also some Aux verbs right-adjoin to main verb to form complex verbs (92ff.); no explicit statement about attachment site of Obj suffixes with these types

8) Definite article in genitive embeddings --- not applicable

--- Genitive construction: Possessor (t_i) Head (36-37); Possessor in non-subordinative case

--- No article; exx. typically show both Nouns with normal pgn suffixes, i.e. not in "indefinite" gender

--- Regarding Dem //nāā, slot sequence of NP specifies order

Dem Possessor t_i Head (21),

thus implicitly allowing Demonstrative to be associated either with Possessor or Head; exx. of this sort (37) realize both possibilities of semantic association; should be impossible to have two Demonstratives

9) Nonconcord of V with full-NP Subj --- not applicable

--- Verb shows no subject concord; Obj nonconcord with full-NP object

10) Verbal abstract: VN or Inf? --- probably INFINITIVE

--- Subordination (121ff.) appears to embed ordinary clauses as such (with ordinary verb rection, per exx. passim), except for deletion of Declarative marker ke; thus for both "nominalization" (126) and "intentional participle" (133-34); additionally, the former may delete NP-subject and tense, rendering it more clearly nonfinite

11) Predicative particle? --- conceivably

--- NomPred may occur either in verbless "equational sentence" (58) or in stative copular sentence (84ff.); both involve Subj-Pred-(Cop) order, with no particle

--- But adverbial ending -se may be added to NPs (replacing pgn ending), thus (99): king ==> king-se "in a kingly manner, as a king"

he king-se is = "He is being (behaving like) a king";

however, this isn't presented as a strategy for forming NomPred per se

12) Postpositional periphrastic (VN + Postp) --- no mention, assume NO
 --- Some Postps can take clausal object (134-35), but in adverbial usage

13) "DO" periphrastic (VN + "DO") --- no mention, assume NO

14) Adverbial clause = "and" + finite clause --- NO (?)

--- Conjunction tsii "and" (117) also functions as adverbial subordinator in a type of "participial construction" (131-32); occurs clause-finally, marks its clause as backgrounded to main predicate:

The prisoners were transported [chains-with bound]-tsii
 "The prisoners were transported bound with chains"

--- Construction deletes both NP-Subj and Tense of tsii-verb, which are construed semantically from main clause (131); thus clause not really finite

15) VN/Inf instead of finite main-clause form --- weakly

--- No well-defined VN; there does exist a special "sequential sentence" construction, involving several conjoined clauses all with same Subj: NP-Subj and ke (Declarative) delete from all but first clause, while tense-aspect-copula complex deletes from all but the last (119-20); in this sense "nonfinite", but not "VN/Inf"

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- apparently not

--- Checked "child", "parent" in Nama Wörterbuch by F. Rust (1969); only "fatherland"

Hua (New Guinea; East-Central Highlands [E. New Guinea Highlands stock])
 Haiman 1980

0) Basic categorial information

--- Noun: case marked suffixally (227ff.), on Erg-Nom basis (363); pronominal possessor marked prefixally or suffixally or both (199, 224, 254, 366ff.); plurality coded optionally in various ways (221ff.); slot sequence (221); no gender; no articles (258ff. on pronominal "definite deictics", not article-like)

--- Verb: object marked prefixally (199-200), sometimes periphrastically; subject agreement expressed through very complex Ablaut patterns affecting (in different ways) both the verb-root and various "mood" desinences (47ff., 61ff.); Medial verbs also take Anticipatory Subj suffixes (57, 375ff.); slot sequence (193)

1) Conjugated adpositions --- category barely present

- Only a few postpositions (271, 336), semantically non-central, and "all of them are nouns" (271); nothing said about pronominal Obj-of-Postp, but (presumably) counts as a pronominal Possessor, which is coded affixally on its Head
- 2) Word order (335ff.)
 - SOV (V-final order rigid in subordinate clauses) (335-36)
 - Postpositions (if category exists at all) (271, 336)
 - Adj-N (268); a few postnominal Adjs (224)
 - Gen-N (257)
 - RC1-N (257, 262ff.)
- 3) Relative clause linker --- ZERO (262-63)
- 4) Relativization strategy/ies --- GAP/COPY
 - Gap coreferential N; change verb-final indicative desinence -e to -ma' (169-70, 262ff.), or change V to its infinitival form (174-75); few restrictions on what can gap (461ff.); possessive RCLs are OK
 - Gapped coreferential N apparently can have pronominal copy left behind, but present even in non-relative clause (384ff.? unclear)
 - RCLs formally identical with N-complement clauses (170); when nominalized, headless RCLs function as variety of adverbial clauses (299ff.), or even "as stylistic variants of simple verb stems" (305ff.)
- 5) Special relative form of verb --- not exactly
 - Ending -ma' said to be specialized for RCLs (169-70); but "RC1" itself also has uses as Adv-Cl and N-complement clause
- 6) Polypersonal verb --- 3 actants coded (unusual combination)
 - Subj (complex Ablaut), Obj (prefixal), Anticipatory Subj (suffixal)
- 7) Infixing/suffixing alternation --- YES (but infixing/prefixing)
 - With verbs embodying the meaningless prefix ha- (36, 196ff.), Obj affixes are infixed, not prefixed (ha-Obj-ROOT) (204); by contrast, no infixing with normal compound verbs (200)
 - Also have empty support-verb to in certain phonological contexts that preclude normal prefixed Obj-marker (202):
 - * Obj-V ==> V Obj-to ;
 creates appearance of infixing, but to acts as a full verb (main V appears in Same-Subj medial form)
- 8) Definite article in genitive embeddings --- not applicable
 - No article; genitive syntagm (366ff.) is apparently
 - Possessor-Gen.case (his)-Head
- 9) Nonconcord of V with full-NP Subj --- NO
 - Full subject agreement (47, cf. 365)
- 10) Verbal abstract: VN or Inf? --- INFINITIVE
 - Infinitive in -di' (56, 174-75); no discussion (?) of rection of Inf, but exx. (317) show Obj in Nominative (as in finite clause); note that Inf does show a minimal two-way coding of Subj via "general ablaut rule" (2/3 person du/pl vs. all other), highly reduced vis-a-vis normal

coding (55-56)

11) Predicative particle? --- sort of

--- Several copular verbs (343ff.), but Cop optional (345); syntagm apparently

Subj NomPred (Cop)

--- Copular verb, like any verb, takes desinence -e (indicative) or -ve (interrog) (248, 62); if Cop omitted, these desinences are instead affixed to NomPred (345), as if a predicative particle; NB: these "orphan desinences" -e, -ve also show up with predicative force on NPs used elliptically instead of full sentences (347-48)

--- In equational sentences (A=B), Cop usually omitted (388), again with -e on NomPred

12) Postpositional periphrastic (VN + Postp) --- barely applicable

--- Few or no postpositions; case-inflected forms of nominalized infinitive occur, but all apparently in various adverbial functions (299ff.)

13) "DO" periphrastic (VN + "DO") --- to a degree

--- Many compound verbs of form "ROOT + hu (DO)" (117ff.)

--- Usually ROOT is not attached to any verbal paradigm; sometimes ROOT identifiable as N, Adj, or Interjection (119-20), or as verb-root borrowed from Pidgin; ROOT can be a nominalized native verb, in which case the "DO" combination has specialized semantics (e.g. repetitive) (121-24); reduplicated verbs are recast as RedupV hu, with hu as the only conjugated element (124ff., 131)

14) Adverbial clause = "and" + finite clause --- not applicable

--- No clausal conjunctions (271); verbal suffix -ve "and/or", but no adverbial semantics indicated (122, 468)

15) VN/Inf instead of finite main-clause form --- in a sense

--- The only plausible candidate would be Medial verbs (clause-chaining), which usually do show full subject coding via Ablaut (187ff.); but "Coordinate Same-subject" Medial verbs show only the two-way opposition of "general ablaut rule" (2/3 person du/pl vs. all other) (58ff.), precisely the same minimal distinction shown by the Inf (recall [10]); thus the form is just as "nonfinite" as the Inf, but is formally quite different from the Inf

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- no data

--- Kin terms (217-18), no mention of metaphorical use

Hungarian (Eastern Europe; Finno-Ugric)

Lotz 1939; also Ackerman 1987

[My thanks to Adam Jacobs for advice and insight on many points]

0) Basic categorial information

--- Noun (56ff.): suffixal inflection for number, case, possessor; no gender; slot-sequence is:

N-Pl-Possessor-Case (63)

--- Verb: suffixal inflection for pers/numb of Subj; Transitive verbs have distinct "Definite" and "Indefinite" conjugations, the former used with definite Objs and partially coding Obj on V (see [6] below)

--- Article: definite and indefinite articles, written as separate words, preposed to Noun (115, 119-20)

1) Conjugated adpositions --- YES, for some Postps

--- Lotz calls Postp by the name "Dependent Adverb" (101ff.)

--- Some Postps take full-NP Obj in caseless form of N; these take pronominal Obj as possessive suffix on Postp (106; AJ, p.c.)

--- Other Postps take full-NP Obj in oblique case-form; these take pronominal Obj as independent Pron (in appropriate case) [AJ, p.c.]

--- A few Postps may take Pron Objs of both kinds

2) Word order

--- Clause-level word order very free, mostly pragmatically controlled by factors like definiteness and focus (253, Ackerman 103ff.) [AJ, p.c.]

--- Postpositions (236-38, 254) (grammar calls them "dependent adverbs"); certain Postps can appear preposed to object, for emphasis (238)

--- Adj-N (230); also marked order N-Adj (appositional) (234-35)

--- Gen has two subtypes (232-33) (see [8] below), one strictly Gen-N, the other with nouns in either order or separated [AJ, p.c.]

--- N-RC1 [AJ, p.c.]; Lotz's discussion unclear (284)

3) Relative clause linker --- Relative pronoun (284-85)

--- Case-marking reflects role of Head within RC1; forms of RelPron (113-15)

4) Relativization strategy/ies --- Gapping, Gap/Copy

--- Little explicit is said; RelPron in clause-initial position (278, 284)

--- Essentially any role relativizable; relativizing on Postposition fronts the Postp along with RelPron; Possessive relativization can only use "Dative" subtype of possession (see [8] below) [AJ, p.c.]:

the boy [who-Dat Peter met [____ the-father-his]]
"the boy whose father Peter met"

--- Verb in RC1 usually in Indefinite conjugation (because RelPron usually counts as an indefinite pronoun) [AJ, p.c.]

--- Subj and Obj RC1s also expressible by participles, per exx. and glosses ("der gehende Mensch, das geschriebene Buch") (164)

5) Special relative form of verb --- NO

6) Polypersonal verb --- Codes 1 and sometimes 2 actants, suffixally
 --- Verb exists in a "definite" and an "indefinite" conjugation;
 Definite conjugation involves implicit 3-person (definite) object (160),
 e.g.

lát-om "I see <3-pers-Obj>" ;

Def paradigm also includes one portmanteau form (125): 1sg ==> 2sg/pl

7) Infixing/suffixing alternation --- somewhat, or NO (depending on analysis)

--- There are compound verbs that have preverbs (102, 105); sometimes Pvb is separable from verb (254-55)

--- A subset of such verbs cannot take an independent Obj pronoun (the usual way of expressing Obj), but must have it "infixing" (Ackerman 82-83, 267, 271):

- (i) neki-megy "attack" (lit. "run into")
 (NB: as a case-marker, -nek = Dat)
- (ii) nek-em-ment "he attacked me"

--- However, another analysis is possible [AJ, p.c.]: when compound verbs take a preverbal full-NP argument, normally the NP is in the appropriate "lative" case and the verb's preverb deletes; (ii) could be construed the same way, as the appropriate (Dative) case-form of Indep pronoun (nek-em), with the preverb (neki-) deleted:

nek-em (neki)-ment ;

here (unlike other compound verbs), preverb and base of Indep pronoun are identical (nek-)

--- In any event, there is no "suffixing" counterpart to this "infixing" Pron

8) Definite article in genitive embeddings

--- 2 basic patterns for possession (233):

- (a) Possessor(Nom) + Head-his: Péter ház-a "P's house"
- (b) Possessor(Dat) + Art + Head-his: Péter-nek a ház-a (SAME);

note absence of article in (a) type, whereas possessed form of Noun normally does take Art:

a ház-am "the-my-house"

--- The (a) type ("short possessive") is very tight linkage; "Dative possessive" (b) is very loose (NPs can switch or be separated [AJ, p.c.])

--- Exx. of both types (233-34) show Possessor NP with and without its own (preceding) Art

9) Nonconcord of V with full-NP Subj --- minimally (252)

--- Only nonconcord (optional) is with collectives or conjoined singular subjects (252)

10) Verbal abstract: VN or Inf? --- BOTH

--- Infinitive in -ni (145-46), little discussion of syntax; Inf takes Obj in Accusative, can occur in EQUI contexts [AJ, p.c.]

--- Also a "verbal noun" in -ás (197); rather like a gerund, Obj in Genitive [AJ, p.c.]

- 11) Predicative particle? --- NO
 --- AdjPred or NomPred is in Nom (with or without Copula) (235-36, 249-50); exx. (251) show order: Subj-Pred-(Cop); no particle mentioned
 --- NB: with verbs like "become", Nom case may be replaced by Dative (84) or Factive (86) case; but not standard with a real "BE" verb (only exception is infinitive of "BE", which can take Dat (84, 236))
- 12) Postpositional periphrastic (VN + Postp) --- trace
 --- Some verbs of change-of-state (lexically restricted) can show construction with Present Participle (not verbal noun) in inessive case (-ban "in") [AJ, p.c.], e.g.:
 spoil-PresPtcpl-ban BE "It is spoiling"
 --- No mention of Postp taking Infinitive as object; although infinitive does not itself inflect for case (145-46), many Postps do govern object in caseless form, so in principle they might govern caseless Inf too
- 13) "DO" periphrastic ("DO" + VN) --- NO [AJ, p.c.]
- 14) Adverbial clause = "and" + finite clause --- NO [AJ, p.c.]
 --- Sentence-level "and" conjunction (273), but no mention of Adverbial use
- 15) VN/Inf instead of finite main-clause form --- not really
 --- Infinitive can occur as bare predicate, expressing "possibility of the action occurring"; but this is elliptical, omitting lehet "it is possible that []" (250)
- 16) Word-initial change --- NO [AJ, p.c.]
- 17) Extended use of kin terms ("Kin of Noun") --- YES, to a degree
 --- Etymological dictionary (Magyar-Angol Szótár) gives numerous examples from fi- "boy, son", (apparently not from other kin terms) [AJ, p.c.]: "son of homeland" = patriot, "son of father" = kinsman, "son of world" = worldly person, "son of church" = sexton, "son of beast" = poultry (or Kleinvieh), "son of fair" = gift bought at fair, "son of chest (or table)" = drawer, "son of window" = glass pane
 --- Some of these occur as compounds, some as explicit genitives, some both; terms tend to be archaic, or verge on archaism

Inghush (Caucasus; Northeast Caucasian)

Nichols 1990 (ms, unpaginated; references by section number)
 [My thanks to Johanna Nichols for answers to several queries]

- 0) Basic categorial information
 --- Noun: gender/number usually covert categories, sometimes overt: if overt, marked on root-initial consonant of Noun (and of Verb or Adj) (2.1.1); plural also marked suffixally (2.1.2); case marked suffixally, Erg-Nom pattern (2.1.3); no article; slot sequence is: N-Pl-Case

--- Verb: about 30% of verb roots (including almost all Aux verbs) code gender/number agreement with verb's Nominative argument (S or O), marked prefixally as on Noun (2.5.2.2); a few verbs show internal consonant change to mark quasi-aspectual plurality (2.5.2.4); in Present, a few verbs distinguish forms with 3-pers Subject (A, S) from non-3-pers forms by internal vowel change (2.5.2.1)

1) Conjugated adpositions --- NO

--- All Prons are independent (2.2, 2.4; JN, p.c.)

2) Word order

--- SOV neutral order (3.3.2), but flexible; OVS/VSO common in main clauses

--- Postpositions (3.3.1)

--- Adj-N (3.3.1)

--- Gen-N (3.3.1)

--- RCl-N (3.3.1)

3) Relative clause linker --- ZERO

4) Relativization strategy/ies --- Gapping (3.5.3)

--- RCl uses Participle (past, present, future, 2.5.2.6.3), which agrees in case with main clause, takes gender/number of Nominative of its own clause

--- Any clause-level function can be relativized; N is gapped with its case marker; relativization on PostpPhr doesn't occur regularly (strands Postp); Possessor relativization restricted to kin and body-part possession

5) Special relative form of verb --- sometimes

--- Past participle also serves as "anterior converb" form (used in clause chaining) (2.5.2.7); other participles have no such secondary usage (?)

6) Polypersonal verb --- sometimes 1 actant, sometimes zero

--- About 30% of verbs (including most Auxes) show gender/number of Nom argument (S, O) (2.5.2.2); a few show separate plural marking of Nom argument; a few distinguish 3-pers from non-3-pers of Subject (S, A); thus have theoretical possibility of coding both Subj and Obj, but would be a "tour de force" [JN, p.c.] (cf. [0] above)

7) Infixing/suffixing alternation --- NO

--- Verbs can have preverbs (2.5.1.1); various elements can intervene between V and Pvb (e.g. WH, NEG), but pronouns cannot; Obj pronoun not a bound morpheme anyway

8) Definite article in genitive embeddings --- not applicable

--- No articles; Possession formed by juxtaposition, with first N in Gen (3.1)

9) Nonconcord of V with full-NP Subj --- NO

--- Those verbs that can show concord do so obligatorily [JN, p.c.]

10) Verbal abstract: VN or Inf? --- "INFINITIVE" (if at all)

--- A variety of "nonfinite" forms (a fuzzy term): "verbal noun", "infinitive", participle (2.5.2.6), converb (2.5.2.7); exx. show Object in Nom case for all of these [JN confirms, p.c.]

--- But NB: criterion for finiteness is presence of tense marking, since all verbal forms (finite or nonfinite) partake in prefixal gender/number coding in same way [JN, p.c.]; these forms thus count as "finite" by criteria used herein

11) Predicative particle? --- NO

--- All copular predicates in Nominative, no particle (3.2.3); syntagm is:

Subj Pred Cop

12) Postpositional periphrastic (VN + Postp) --- not applicable

--- Various verbal periphrastic constructions (2.5.4), some with copula (2.5.2.3), but none involving Postp.s or oblique case forms; no "Infinitive" (see [10])

13) "DO" periphrastic (VN + "DO") --- in a sense

--- New verbs can be coined only periphrastically, with "DO" preceded by a borrowed root (2.5.3.2); also occurs with some roots that "are not obviously borrowed"; these roots aren't connected to a verbal paradigm, do not fall into any standard part of speech

14) Adverbial clause = "and" + finite clause --- not applicable

--- No "and" conjunction (3.4)

15) VN/Inf instead of finite main-clause form --- not applicable

--- Clause chaining makes use of two special "converb" forms, marked only for Anterior vs. Simultaneous aspect (3.6), but not for tense in the usual sense; nonfinite, but have no uses as "VN/Inf"; in any event, these count as "finite" by standards used herein

16) Word-initial change --- not really

--- Gender/number marking and agreement involves a prefixal alternation (2.1.1); but not a phonological process

17) Extended use of kin terms ("Kin of Noun") --- apparently not [JN, p.c.]

Japanese (East Asia; isolate)

Hinds 1986; Matsumoto 1988, Hasegawa 1992

[My thanks to Yoko Hasegawa for advice and insight on many points]

0) Basic categorial information

--- Noun: no gender, optional plural suffix -tachi (171, 222ff.); "case" coded by postpositions (187); no article (82)

--- Verb: no person/number coding (323, 327)

- 1) Conjugated adpositions --- NO (354)
- 2) Word order
 - SOV (Intro iii)
 - Postpositions (187, 350)
 - Adj-N (82)
 - Gen-N (82)
 - RCl-N (61, 82), also internally headed (62)
- 3) Relative clause linker --- ZERO (63)
- 4) Relativization strategy/ies --- GAPPING
 - RClS are normal finite clauses, precede HeadN, no special mark (58-59); coref NP standardly gaps (61-62) (and its Postp gaps with it: Matsumoto 32); but sometimes can have pronominal copy, e.g. if Postp is preserved or if have possessor relative (61-62)
 - Subject of RCl may be marked with Gen Postp rather than Nom (53-55, 60)
 - Also internally headed type (62)
- 5) Special relative form of verb --- barely
 - Special RCl-form only for non-past copula (59)
- 6) Polypersonal verb --- zero actants
- 7) Infixing/suffixing alternation --- not applicable
- 8) Definite article in genitive embeddings --- not applicable
 - No article; genitive syntagm is: Gen no Head
- 9) Nonconcord of V with full-NP Subj --- not applicable
- 10) Verbal abstract: VN or Inf? --- INFINITIVE
 - Nonfinite forms defined as lacking tense (323); such forms show same rection as finite verbs (199)
- 11) Predicative particle? --- NO
 - Syntagm (69-70) is: Subj Pred (Cop), with no particle
- 12) Postpositional periphrastic (VN + Postp) --- sort of
 - Can productively form construction: VStem tutu BE, meaning "be in the process of VERBing" or "about to VERB"; tutu is a particle that can also function as clause-connective ("while"), but otherwise no other uses [YH, p.c.]
 - From verbs of the form "VN + DO" (see [13] below), can productively form construction: VN-tyuu COP, meaning "be engaged in VERBing"; -tyuu not a postposition but a derivational element meaning "in, middle", unattested as free form but written with same character as words for "center"; however, this is not a true "Verbal Noun" [YH, p.c.]
 - Cf. also construction of Perfect (299-300) and Progressive (303-4): V-te BE, with the special nonfinite "te-form" of the V; but no Postposition involved
- 13) "DO" periphrastic (VN + "DO") --- in a sense

--- Productive process creates verbs of the form: VN + suru ("do") (371-72); element that precedes the verb is typically a foreign word (Chinese-based or otherwise; often a foreign verb), or an Adv or ideophone [YH, p.c.], but almost never a true verbal-noun from a Japanese verb (rare counterexample: hanasi suru vis-a-vis hanasu, both "speak" [YH, p.c.]

14) Adverbial clause = "and" + finite clause --- unclear; conceivably --- No straightforward clause-level "and" conjunction (84ff.); sentence-level conjunction ga is usually translated "but" (89ff.), but sometimes better "and" [YH, p.c.]; however, no AdvCl use; AdvCl normally formed as RCl, or with adverbial conjunctions, or with special nonfinite te-form (63ff.)

--- Clause-level conjunction to, glossed "if" in grammar (67), is sometimes better rendered "and" [YH, p.c.]; if so, "and"-clauses and "if"-clauses (adverbial) overlap; note also NP-level conjunction to "and" (93ff.), possibly not merely homophony [? my suggestion]

15) VN/Inf instead of finite main-clause form --- YES

--- Nonfinite te-form regularly used in clause chaining (84) and some serial verbs (?) (331); this form does have uses corresponding to prototypical "VN/Inf" uses, viz. Object of Aux for numerous Aux (299-332 passim; see also Hasegawa 1992)

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- NO [YH, p.c.]

Kiowa (Southern Great Plains; Kiowa-Tanoan)
Watkins 1984

0) Basic categorial information

--- Noun: number marked suffixally, in a special way: nouns lexically specified as either "inherently" sg/du or "inherently" du/pl, and an "inverse" marker toggles this number specification, converting sg/du nouns to plural, and du/pl nouns to singular (78); also possessive prefixes for some nouns (see [8] below) (101); apparently no cases ("semantic role is unmarked on fully specified nouns": 204), but various "locative-directional" suffixes (see [1] below) (53, 188) are perhaps somewhat like case endings

--- Verb: very complex system of prefixes coding not syntactic but semantic categories (109-10): markers conflate Agent or "Patient" (human; idiosyncratic usage) and inanimate "Object" (110-14); also, verbs of position occur in number-suppletive pairs, sg/du vs. pl (153-54), as also do stative adjectivals (sg vs. du/pl, 154); verbs freely incorporate nouns, verbs, and adverbials (224ff.)

--- Article: none discussed; there is a discourse suffix glossed "definite" (191-92), but it apparently attaches only to adverbial deictics, not to nouns in general (186-87)

- 1) Conjugated adpositions
 - No category "adposition" mentioned in grammar; instead, have "locative roots" (up, down, out, beyond) (193), which take various locative/directional suffixes (table, 188)
 - There are also other, different loc/dir suffixes which attach only to ordinary nouns (193); these are nowhere listed systematically; exx. are: "at" (53), "on" (230), "to" (235), LOC (254), INSTR (211); syntax sketched, p. 210
 - Categorical status of these endings unclear; not presented as case endings or as postpositions; but not really discussed at all, or presented as a unified set
- 2) Word order
 - Agt-Pat-Obj-Verb order basic (205); but old information easily dislocatable to right of verb (206)
 - No category Adposition
 - No category Adjective; adjectival modification either by RC1 (with stative verb) or by compounding (99, 208)
 - Gen-N (107)
 - RC1: analyzed as internally headed (232); sometimes HeadN may be preposed to entire RC1 (233), thereby yielding order N-RC1
- 3) Relative clause linker --- not applicable
 - Internally headed
- 4) Relativization strategy/ies
 - RC1 analyzed as internally headed; looks like any normal clause, except that final word in RC1 is suffixed with a number affix (either "basic" -dè or "inverse" -gò, depending on class and number of HeadN) (230); a locative/directional suffix may follow -dè/-gò (234)
 - Note that final word in RC1 need not be the verb (per ex., 206)
 - RC1 may be optionally preceded by subordinating particle Ógò (231), which may be optionally followed by relative-anaphoric particle ám "that" if HeadN is definite or previously mentioned
 - Overall RC1 syntagm:
(Ógò) (ám) [... HeadN ...] -dè/gò (Loc/Dir)
 - In all (?) exx. cited, HeadN is at extreme left edge of RC1; but position of Ógò supports analysis as internally-headed clause
 - Other options: Head may be preposed to RC1 (i.e., preceding Ógò) (233); or entire RC1 may be moved to right of matrix-clause verb, with the HeadN either staying inside the RC1 or left behind in matrix clause; these subtypes seem to involve Gapping
 - Note that HeadN does not have to be coindexed on V (as if Gap/Copy); (per ex., 234)
 - No information on possessive relativization; presumably works unproblematically in internally-headed type
- 5) Special relative form of verb --- NO
 - Particle -dè/-gò not constrained to appear specifically on the verb (ex., 206)
- 6) Polypersonal verb --- 2 (or 3) actants coded
 - Prefixal markers conflating Agent, Patient (human), Object (inan), according to 4 paradigms (111):

- a) Intransitive --- 1 role coded
- b) Agent/Object --- 2 roles (Agt, Obj) coded
- c) Patient/Object --- 2 roles (Pat, Obj) plus implicit Agt
- d) Mixed/Object --- 2 roles (Agt, Obj) plus implicit Pat

--- The prefixes typically do convey nonvacuous pers/numb information about the "implicit" arguments (table, 116)

7) Infixing/suffixing alternation --- NO

--- All markers prefixal, conflated

8) Definite article in genitive embeddings --- not applicable

--- No definite article

--- Genitive syntagm depends on type of HeadN (101): body parts, kin terms, other; patterns are:

- (a) Body parts: no possessor affixes (possessor marked on V)
- (b) Kin terms: possessor affixes
- (c) Other: possessor forms a compound with HeadN

--- Possessor affixes ("my, your", etc.) are prefixes, but 3-sg possessor also takes suffix -d/té (102)

--- Noun-Noun possession (including of kin terms) done by forming a compound: Possessor + Head (107); but if Possessor is proper name, must again use -té:

Laurel-té cègùn "Laurel's dog"

9) Nonconcord of V with full-NP Subj --- NO

--- To the contrary: ambiguity in Noun number (sg/du or du/pl) is resolved when the N is crossreferenced on the V, since V markers do distinguish sg, du, pl (79, 112); Verb thus has a richer coding of number than does Noun

10) Verbal abstract: VN or Inf? --- neither

--- No discussion of any verbal abstract; all clauses (including complement clauses, 235) appear to be finite; verb pers/numb prefix obligatory (109)

--- Many EQUI constructions are handled by incorporating subordinate verb into matrix verb ("Raising Incorporation", 228-29); if subordinate verb has an object, it too is raised and incorporated:

I/him - meat+buy+send - perf = "I sent him to buy meat"

--- "Nominalization" mentioned in passing (61, 63), not discussed (?); probably a type of Noun+Verb compounding (75-78); all exx. seem to involve concrete objects, not verbal abstracts

11) Predicative particle? --- NO

--- There is a copular verb -dó: "be" (227); the normal way to handle predicate nominals is

Subj Pred BE (exx. 105-6, 227; with all three elements, 231)

--- May also incorporate Pred into verb BE; changes meaning: "to be Pred-like"

12) Prepositional periphrastic (Prep + VN) --- not applicable

- No category adposition
- The verb "be" forms incorporated compounds with other verb stems to form statives (151-52); but no Adp, semantics not right
- 13) "DO" periphrastic (VN + "DO") --- puzzling
 - Verb roots can be incorporated into the root tó: "behave, act", the latter acting as an auxiliary so as to yield imperfective form of main verb (156); imperfective covers the nuance of "events in progress" (158), i.e. progressive; normally "the auxiliaries do not contribute additional lexical content" to the root, though there can be change in nuance (156-57); note that verbs also have a synthetic, inflectional imperfective (157ff.)
 - This is a good match for [12] semantically but not syntactically; it's a rough match for [13] syntactically but not semantically (Aux verb is not "do" but "act")
- 14) Adverbial clause = "and" + finite clause --- perhaps (semantics)
 - Clauses usually coordinated via switch-reference (236), which involves finite clauses; one switch-ref marker is glossed "and, if (neutral/sequential/conditional)" (236); OK if we admit "if" as adverbial
- 15) VN/Inf instead of finite main-clause form --- not applicable
 - No nonfinite forms; switch-reference forms (clause chaining) are fully finite verbs (per exx. passim) (236ff.)
- 16) Word-initial change --- apparently NO
 - There is initial ablaut of verb roots, a striking feature of Kiowa (and Tanoan in general) (60); this involves hard and soft variants of consonants, morphophonemically conditioned (e.g. in verb incorporation); but apparently these never involve word-initial change, since verb prefixes are obligatory (109) (even in Imperatives: exx. 169)
- 17) Extended use of kin terms ("Kin of Noun") --- no data

Koasati (Southeastern US; Muskogean)

Kimball 1991

[My thanks to Pam Munro for clarification of several points]

0) Basic categorial information

- Noun: 7 cases (388), several articles; both categories suffixal, and are (always?) attached only to the last word of the NP (480); no gender mentioned; plural (rare) only for humans, suffix -ha (403)
- Verb: affixal marking of Subj, DObj, IObj person/number (slot sequence, 113); plurality in 3rd person not part of standard pers/numb paradigm (57), but is marked in a variety of ways elsewhere on verb (136-38, 314ff.); not clear if DObj and IObj can cooccur (since 3-person DObj = \emptyset), but two IObj can cooccur (133)
- Article: suffixal articles (= previous mention + temporality)

(404ff.); much less common in texts than in elicited data (405)

1) Conjugated adpositions --- YES

--- There are locational postpositions (495), though not formally distinct from nouns; used equally as postpositions and as bare locational adverbs; analyze here as "hedged postpositions"

--- When Obj of Postp is a full-NP, it is in zero "Autonomous" case (395); pronominal Obj shows up as possessive prefix (496); but (unlike true possessive syntagm, see [8]), full-NP Obj is not doubled by pronominal mark on Head (per exx.)

2) Word order

--- SOV (513)

--- Postpositions (if a distinct category) (495)

--- N-Adj (though "Adj" characterized as nominalized V) (479); N is in zero-marked "Autonomous" case when followed by Adj (395-97)

--- Gen-N (442)

--- Unclear; grammar asserts N-RC1 (288), but for most exx. (288ff., 525-26) analysis as internally headed RC1s fits much better

3) Relative clause linker --- ZERO

--- Analysis as internally-headed RC1 (below) would preclude such a linker

4) Relativization strategy/ies

--- "Koasati does not have any distinguishing features that mark certain clauses as being relative" (525); common equivalents are two independent clauses linked by switch reference (526), or use of participles (525, 288ff.), or (especially for locational RC1s) use of the nominalization Kimball misnames "Agentive nouns" (282)

--- Grammar presents participial strategy as External-Head followed by modifying participle, with case-mark of Head stripped away just as in N-Adj syntagm (288); but these "participial" verbs have Subj and Obj markers, and are accompanied by full-NP participants; coreferential noun takes case marking appropriate to embedded clause, not matrix clause (per exx. on p. 289, 525); one example (NB) cannot be externally headed (525):

[my-father-NOM woman saw]-NOM your-wife is
"The woman that my father saw is your wife"

--- Thus participial strategy appears to be internally headed; indeed, these "participles" don't much resemble prototypical participles

--- Examples of participial strategy include Subj, Obj, and locational relativization; no mention of relativization on possessives; for locational RC1s, "Agentive N" strategy favored; no reason (?) to view locational relativization as relativization on Obj-of-Postp

5) Special relative form of verb --- YES

--- Participles (coded by special suffixes) appear to be specialized for RC1 function, and are used "exactly like noun modifiers (adjectives)" (288)

6) Polypersonal verb --- 3 affixal actants coded (56ff., 113, 127ff.)

7) Infixing/suffixing alternation --- NO

--- DirObj and IndObj have fixed place in sequence of "rigidly fixed slots" (111)

--- There exist Aux verbs (89ff.), in syntagm "MainV - Aux"; only Aux inflects for Subj, with MainV usually chained to it with connective suffix -t (per exx., see [15] below); but MainV apparently can take Obj suffixes as usual (exx. 94-95 show MainV marked with IndObj; no DObj exx.?)

8) Definite article in genitive embeddings

--- Possessive syntagm: Gen-HeadN; full-NP possessor is doubled by pronominal possessor prefix (often zero) on HeadN; no explicit discussion of article placement in this syntagm, but recall that articles aren't very common anyway; in purely pronominal possession, article and possessor prefix may cooccur (per ex., p. 410: "my-grandfather-Art")

--- Find exx. (one apiece) of article on either Head or Dept:

- | | | |
|---------------|--------------------|--|
| (i) my-mother | her-father-Art | "my mother's father" (442) |
| (ii) aunt-Art | her-brother.in.law | "this aforementioned aunt's
brother-in-law" (405-6) |

9) Nonconcord of V with full-NP Subj --- apparently NO

--- Nouns seldom code plural (403), and then only optionally; rather, the norm is for noun plurality to be rendered indirectly, via verb plurality (446) (i.e., the exact reverse of the nonconcord at issue here)

--- Recall that verb plurality not coded on Subj/Obj prefixes, but elsewhere on verb

--- Exx. (403-4, 280) appear (?) to show an explicitly plural-marked Subj noun taking a plural-marked Verb

--- Concord of conjoined NPs and counted NPs unclear

10) Verbal abstract: VN or Inf? --- apparently Inf

--- Grammar discusses "verbal nouns", which "serve functions similar to the English infinitive or gerund" (273); but no discussion of rection of object; distinguishing accusative from genitive rection should be possible, but difficult in practice

--- With pronominal (affixal) arguments, DObj affixes (113) are identical in form to inalienable possessor affixes (432); however, alienable possessor affixes differ (432); examples with pronominal object of verb thus ought to decide the question unambiguously; but I found none

--- With full-NP arguments, DObj would take Acc case (-n, 392-94), while Possessor would take "Autonomous" case (zero ending, 395); on the other hand, Acc case itself has a frequent zero allomorph (394); still, the syntagms should differ:

DObj Verb	vs.	Possessor-Ø its-VN ,
-----------	-----	----------------------

where "its" (Possessor affix) would be nonzero for alienable possessor; the few exx. of "verbal noun" with full-NP object (275-81) appear to show Obj in zero-form Acc, but no "its" affix, arguing for Infinitival status

11) Predicative particle? --- NO

--- Predicate nominals expressed with Copula (99, 101-2), which is omissible (501); but no hint of any particle (nor in exx.)

- 12) Postpositional periphrastic (VN + Postp) --- no mention, assume NO
 --- Recall that Postp is "hedged"
- 13) "DO" periphrastic (VN + "DO") --- no mention, assume NO
 --- List of "Auxiliary" verbs (89ff.) include no "DO" verb
- 14) Adverbial clause = "and" + finite clause --- not applicable (?)
 --- List of "conjunctive words" (536-39), which link clauses and "relate the actions of the previous sentence to the following sentence in terms of sequence, dependence, and causality" (536); but none is simply glossed "and"; construction cooccurs with switch-reference marking, would appear to involve clause-chaining; no sentence exx. given; see also [15] below
- 15) VN/Inf instead of finite main-clause form --- weakly
 --- Verbal conjunction done with clause-chaining connective suffix -t (roughly simultaneous action) (527), where only the final such linked verb is person-marked (227) or can take verbal suffixes; contrast switch-reference marker -k (same subject), which does allow these markers on non-final verbs (227, ex. on 528)
 --- This -t form is thus not finite; but not a "VN/Inf", and normal "verbal noun" is not used in this way (278)
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- no data

Kobon (New Guinea; Kalam family [East New Guinea Highlands stock])
 Davies 1989

- 0) Basic categorial information
 --- Noun: no case (147), no gender (149); number and possessor not coded except on kin terms (60, 147), sometimes via suppletive stems; nouns are very frequently doubled by following (independent) pronoun (151, 157); personal pronouns have special Object case-form (159), but its use is optional
 --- Verb: only Subj is marked, suffixally (183-84; paradigms 168ff.); heavy use of serial verbs (203-4) and clause chaining with switch reference (184ff.)
 --- Article: indefinite article only (postnominal particle) (60, 150)
- 1) Conjugated adpositions --- NO
 --- Neither the postpositions (there are four, 107, 205-6) nor the various Postpos-like locative words (like "outside", 206) enter into combination with pronouns (206); Kobon has no reduced pronouns (152)
- 2) Word order
 --- SOV, fairly rigid (47, 107)
 --- Postpositions (107)

- N-Adj (57)
 - Gen-N for Noun possessor, but N-Gen for (independent) Pronoun possessor (57, 113); several exx. of doubly marked syntagm (31, 157):
GenN HeadN GenPron (boy bird he = "the boy's bird", 31)
 - RCl-N, or internally headed (28ff., 58)
- 3) Relative clause linker --- apparently ZERO
- Internally headed RCl may be immediately followed by resumptive pronoun copy, agreeing in number with Head N (30); perhaps this is a "linker", but seems more like a biclausal, co-relative strategy ("Which N ..., THat ...") (though not presented as such)
- 4) Relativization strategy/ies (28ff.) --- Gapping, or full-NP-Intact
- Order is RCl-N, or internally headed (28); if externally headed, HeadN may also recur within RCl (NB: no mention of pronoun copy); any occurrence of coreferential N (as head and/or internally) may optionally be followed by Demonstrative Pronoun (28-29); in internally headed RCl, RCl may be followed by a resumptive pronoun (30)
 - Can relativize on any embedded role (31-32); in externally headed RCl, gapping is preferred, but can retain full-NP intact for clarity or disambiguation (29)
 - Subject-relativization also possible via an "Adjectivalizing" verb-suffix -ep/-eb (31), also used as Nominalizer (26-27); -ep/-eb attaches to verb stem, without subject inflection
- 5) Special relative form of verb --- minimally, for one subtype
- Suffix -ep/-eb not limited just to RCl function (also nominalizer)
- 6) Polypersonal verb --- 1 actant coded (183-84)
- 7) Infixing/suffixing alternation --- not applicable
- No preverbs, nor any suffixal analogue; Obj pronoun not an affix; grammar has one example where Obj pronoun precedes a V+V serial-verb chain (165)
- 8) Definite article in genitive embeddings --- not applicable
- No definite article; no data provided for indefinite article
- 9) Nonconcord of V with full-NP Subj --- NO
- The only deviancy from full concord is that a special set of kin terms (even when semantically singular) demands plural concord (and plural pronominal reference) (148-49, 153); clearly a special case
 - With conjoined Subj NPs ("X with Y"), verb may show concord with NP as a whole, or just with X (73)
- 10) Verbal abstract: VN or Inf? --- INFINITIVE
- Several nonfinite forms lack person-marking (182-83):
 - i. "Infinitive", used in serial verb constructions (203-4)
 - ii. Purpose marker -niq (37, 183)
 - iii. Simultaneity marker -öl (36-37, 183)
 - Genitival (Gen-N) and Obj-V syntagms are formally distinguishable only for Pronoun dependents, because (a) pronouns show order N-Gen, and

(b) pronouns do show a distinct Object case; hence no reliable way to tell apart VN from Infinitive except with pronominal Obj; one ex. (165) shows explicitly non-genitive construction with pronoun; no hint elsewhere of nonfinite verb forms taking genitival rection of Obj, hence assume Infinitive (not VN)

--- Very little discussion of Object complement-clauses (a standard EQUI context); nominalizer -ep/-eb (26-27) can used here, same issues as just discussed

11) Predicative particle? --- maybe (AdjPred)

--- Ordering of PredN and PredAdj is: Subj Pred (Copula) (47); copula optional (42-44)

--- NounPred takes no particle (26, 42-44)

--- Adj cannot occur as "bare predicate" (42-43); various techniques, two involving suffixed particle rö "like" or bö "being something" (42, 204, 211-12):

he	big	<u>rö</u>	"he is quite big"
sky	red	<u>bö</u>	"the sky is red"

rö is basically a postposed comparative particle "like" (93), (though not listed among postpositions, 205-6); unclear whether its use with AdjPred adds nuance "like", or just effects predication

12) Postpositional periphrastic (VN + Postp) --- no mention, assume NO

13) "DO" periphrastic (VN + "DO") --- YES

--- 3 periphrastic constructions involving nonfinite V + gĩ "do" :

i. "Inf" + gĩ: used productively to incorporate foreign verbs (203)

--- ex.: selim gĩ "do selling" (from Tok Pisin)

--- note frequent use of: Ideophone + gĩ "make a noise" (231-32)

ii. V-nĩg + gĩ: Prospective aspect ("be ready/going to V") (175, 183)

iii. V-öl + gĩ: Simultaneous aspect (w.r.t. following clause) (174)

14) Adverbial clause = "and" + finite clause --- apparently NO

--- AdvCl formed chiefly by asyndetic juxtaposition or clause chaining (33-40); neutral "and" expressed by forms of the verb "perceive" [sic!], "which occurs as intersentential or interclausal conjunction" (67)

15) VN/Inf instead of finite main-clause form --- weakly

--- "Inf" may occasionally replace Same-Subject medial forms (which do code Subj) in switch-reference clause chaining (183); no clear distinction between serial verb constructions (203) and such clause chaining (183); in serial verbs, all verbs except the last are again "Inf" (182-83, 203-4); this "Inf" form, in any event, has no prototypical uses of a true VN/Inf

16) Word-initial change --- NO

--- Kobon has hardly any morphophonology at all (228-29)

17) Extended use of kin terms ("Kin of Noun") --- seemingly not

--- No examples readily apparent in "Lexicon" section (233-45)

Lahu (Southeast Asia; Lolo-Burmese [Sino-Tibetan])

Matisoff 1973

[My thanks to James Matisoff for advice and clarification]

0) Basic categorial information: Isolating

--- Noun: no cases, except (optional) postpositive Obj-marker thà? (155); plural suffix -hì used only sparingly with common nouns (65); numeral classifiers (88ff.)

--- Verb: no actants coded; much use of concatenated verbs to comprise a single verb nucleus (= serial verbs) (199)

--- Article: no "article" presented in grammar; but postnominal determiner chi "this" (which can also appear autonomously, 111) often functions like "the" in anaphoric usage ("previous mention") (112); chi has much broader privileges of occurrence than other demonstratives, is introduced by a special phrase-structure rule (111-12); will analyze as "hedged article"

1) Conjugated adpositions --- NO

--- No form-class "postposition" presented per se, but class of "Noun particles" seems to fill the role (153ff.); pronouns are presented only as independent words (49-50)

2) Word order

--- V-final (39-40, 192); usually SOV (191)

--- Postpositions (47); no such form-class explicitly presented, but equate to "Noun particle" (153ff.)

--- No form-class "Adj" presented (44-45); rather, adjectival verbs (193); thus adnominal Adj is really adnominal RC1; either order quite common (491)

--- Gen-N (141), also N-Gen (150ff.)

--- RC1-N normal (473), also N-RC1 (490ff.)

3) Relative clause linker --- Rel-Particle

--- Marker -ve (472), actually an all-purpose subordinator (360); sometimes deletable (487ff.)

4) Relativization strategy/ies --- GAPPING, apparently

--- "No RC may contain a NP that is co-referential with the [Relative Head]" (473); discusses non-occurrence of Subj and Obj NPs in RC1

--- Normal word order is RC1-HeadN (473), with linker ve:

RC1 ve HeadN ;

can postpose [RC1 ve] sometimes, especially if RC1 consists only of a bare adjectival verb (194, 491); [RC1 ve] can also be postposed to end of sentence (511)

--- No discussion of Possessor RC1s, but note ex (479):

[roof NEG high]-ve house "house whose roof is not high"; construction appears restricted to the type "house [high (its) roof]"; probably not relativizing on Possessor NP but on Topic NP in embedded clause (underlyingly: [House(TOP), roof (is) high]); cannot (?) relativize on possessors in other syntactic contexts [JM, p.c.]; will not count this as genitival RC1

--- Also examples of RC1 with spatio-temporal Head (gapped in RC1), e.g. a word meaning "before" (477, 487-88)

--- No discussion of Postp RCl, apparently impossible [JM, p.c.]; important to note that it's very common for oblique NPs to occur (even in non-relative clauses) with no special functional marker (306ff.)

5) Special relative form of verb --- NO

6) Polypersonal verb --- zero actants

7) Infixing/suffixing alternation --- no, or not applicable

--- All NPs (including pronominal) precede verb-complex (40); so no chance of pronominal Obj penetrating into verb-complex even if Obj were taken as clitic; but NP can permute with entire verb-complex (506ff.)

8) Definite article in genitive embeddings --- hedged

--- Possessive syntagm is: Possessor ve Head (141); particle ve omissible if Possessor is a pronoun, and elsewhere (147-49)

--- Explicit statement (144) that "article" chi may appear on Head or Possessor or both

--- Also can have: Head [Possessor ve] (150ff., 510)

9) Nonconcord of V with full-NP Subj --- not applicable

--- Note too that plural is not a verb category

10) Verbal abstract: VN or Inf? --- not applicable

--- No verbal abstracts per se; many different nominalizing particles which go after full clauses (440ff.)

11) Predicative particle? --- sort of

--- Two constructions, of general form Subj Pred (Copula):

(a) With copular verb: phE? "be a certain way/thing" (231-32, 278, 658), or chE "be in a place" (654); no special mark on Pred

(b) Nominal sentence (verbless), termed "minor sentence" (40); here Pred can be followed by declarative (367) or interrogative (372, 374) "final unrestricted particle", e.g. declarative yò (367), which however also occurs freely after verbs, hence not a special predicative particle

12) Postpositional periphrastic (VN + Postp) --- in a way

--- "Noun particle" definitionally can occur only after Noun, so equating "Postposition" with "Noun particle" (see [1] above) excludes Postp from occurring after a Verb (45, 47); apparently no lexical overlap of "Noun particles" (155), "Verb particles" (317, 331, 335), "Unrestricted particles" (390)

--- But: locative copula chE "be in a place" can follow main verb as helping verb ("versatile verb"), in the meaning "continuative" (237, 240), e.g. changes "put on, wear" to "is wearing"; syntax/semantics is almost right (use of locative to express continuative), despite lack of adposition

13) "DO" periphrastic (VN + "DO") --- YES (but unusual semantics)

--- Verb te "do, make" can occur before action verbs in the meaning "make (something) and VERB with it" (244, 213, 552 [note 41]); this usage has nothing to do with causation (244); parallel to similar usage with verb yù "to take", yielding "take (something) and VERB it" (213)

--- Additionally, te is causativizer and transitivizer (244-45)

14) Adverbial clause = "and" + finite clause --- maybe

--- Conjunction le "and" (sentence-level 397, NP-level 177), called the "Suspensive" (417); various English translations explicitly given for syntagm VERB1 le VERB2 (417), including (inter alia):

VERB1 and; having VERB1-ed; while VERB1-ing; after VERB1-ing
 --- VERB1 (though uninflected) is less than fully "finite", because only the last verb in a chain can take "final unrestricted Particles" (P-uf), conveying speech act force, etc. (365-66); JM feels this construction to be clause-chaining (like Japanese) [p.c.]; still, in final finite clauses P-uf is not obligatory (and often absent), so typically the clause preceding "and" could indeed stand alone as a finite clause, which would be the desired construction

--- Also a homophonous le "Causal, because" (408); grammar clearly distinguishes the two syntactically, but says they're often hard to tell apart (418); probably historically related (591 [note 29])

--- Section devoted to "Adverbial expressions" characterizes Adv expressions as being subordinate to a following verbal nucleus (265); but the above syntagms might still be Adverbial equivalents

15) VN/Inf instead of finite main-clause form --- not applicable

16) Word-initial change --- a bit

--- Find a handful of lexical pairs as fossils of two once-productive alternations involving pre-Lahu initial nonglottalized vs. glottalized initial segments (the latter from earlier *s- [520, n. 48]); presented (29-34) primarily as a matter of tonal variation, but some pairs manifest this in part as voiced/voiceless variation in initial segment; viz.

(i) Simplex vs. Causative (32-33)

(ii) Adj ("long") vs. Neutral extensive ("this length") (17-18, 31)

--- In both (i) and (ii), the second member of the opposition had original initial *?-, and is synchronically voiceless

17) Extended use of kin terms ("Kin of Noun") --- some

--- A few "mother" exx. mentioned in Matisoff 1992:35 (and 1988:959 [Dict]), though (NB) not using the normal word for "mother": magnet = iron-mother, thumb = hand-mother, big toe = foot-mother, alphabet = letter(s)-mother, capital investment = money-mother; also pupil (eye) = eye-daughter (1988:1258); nothing further under "father", "mother", "son", "daughter"

Lango (East Africa; Western Nilotic [Eastern Sudanic])
Noonan 1992

0) Basic categorial information

--- Noun: no case (119), no indication of gender; plural only coded on some nouns (human, animals, tools), usually suffixally (83-85, 166-67); possessor suffixes, partially distinguishing alienable/inalienable (77-78); order is: Root-Plural-Possessor (85)

--- Verb: two arguments (141): Subj prefixally (91, 136), Obj suffixally (141-42)

--- Article: no real definite article; have suffix -mÉrÊ "the aforementioned"; indefinite suffix -mÓrÔ, almost always just for Subj; Subj normally taken as definite unless -mÓrÔ appears; with non-Subj, definiteness normally determined only pragmatically (no marking) (161-62)

1) Conjugated adpositions --- YES

--- Of the 9-10 true Prepositions listed, almost all take pronominal inflections (107, 170); identical to inalienable possessor suffixes (107)

2) Word order

--- SVO unmarked, order fairly rigid (119)

--- Prepositions (107)

--- N-Adj (154); HeadN always first in NPs

--- N-Gen (154)

--- N-RC1 (154)

3) Relative clause linker --- INVARIANT

--- Invariant linker àmÊ (215); contains the general attributive particle à that links all modifiers to Noun (154); optional, may be replaced by à or deleted entirely (217)

--- àmÊ has a few headless adverbial-clause uses (243-46)

4) Relativization strategy/ies --- Gapping, Copying

--- Syntagm: HeadN àmÊ RC1 (215); RC1 follows other Noun modifiers (154)

--- Gapping for Subj, Obj RC1s; Copying for genitival and Prep RC1s (215-16)

--- WH-questions and clefting formed by RC1 (219-20)

5) Special relative form of verb --- under restricted circumstances

--- In Subj RC1s (218), in perfective aspect only (not habitual, progressive), 3sg verbs have special tone contour, such that 3sg-Rel = 3pl (92-93, 136-37);

--- NB: this same special form also used in independent (non-relative) clauses if explicit 3sg Subj pronoun occurs (vs. full-NP subject or zero subject) (137)

6) Polypersonal verb --- 2 actants coded (Subj, Obj)

7) Infixing/suffixing alternation --- NO (Obj only suffixal)

- 8) Definite article in genitive embeddings --- not applicable
--- No definite article
--- Genitive syntagm is: (a) Alienable: Head à Gen
(b) Inalienable: Head Gen (156-57)
--- No exx. given with determiner suffixes, which normally should occur on last word of NP (155)
- 9) Nonconcord of V with full-NP Subj --- YES
--- Full-NP Subj always takes 3sg verb (167-68, cf. 137); 3pl verb-form may only take Indep 3pl pronoun as its explicit Subj
--- A few Adjs have special plural stem when used as PredAdj with plural Subj; but such an Adj (which inflects like a verb) takes 3sg prefix (104, 147)
- 10) Verbal abstract: VN or Inf? --- INF
--- Two types of nonfinities mentioned: Inf, Gerund
--- Infinitives take same Obj marking as do finite verbs (141, 213); used in EQUI contexts, nominalizations (213)
--- Gerunds (morphology, 102; a special Niloticist sense of the term, 175) used only in a few contexts:
(a) Type "a book for reading" (214)
(b) Type "Chicken is good to eat" (214-15)
(c) Reduplicative construction, to lend mild emphasis; Gerund follows finite verb and its complements, no indication of it taking an Obj of its own; e.g.: man [killed dog] kill-Gerund (175)
--- Gerund thus apparently never takes syntactic Obj, hence question of Obj rection irrelevant
- 11) Predicative particle? --- NO
--- Syntagm: either (a) Subj PredN/Adj (144, 146); here PredN/Adj takes verb Subj prefixes (144)
or: (b) Subj Cop PredN/Adj (145)
using verb bèdò "stay, sit" as Cop to indicate non-present time
--- Noun Subj may take Indep pronoun copy right after itself
--- "Affirmative particle" dó may appear with such sentences (145), e.g.
òkéìlò dó "It's Okelo";
but also found with normal verbal sentences, conveying "friendly assurance" (186-87), hence not a "predicative particle"
- 12) Prepositional periphrastic (Prep + VN) --- NO
--- In addition to normal (synthetic) Progressive aspect, can form a periphrastic Progr with the verb tíê "be present" as Aux:
tíê-Habitual Verb-Progr (139-40);
but here main verb is finite, construction is paratactic
--- NB: synthetic Progr was originally ya- "be in a place" plus Inf (91)
- 13) "DO" periphrastic ("DO" + VN) --- apparently not
--- No mention of "do" (tímò) in list of Aux verbs (140)

- 14) Adverbial clause = "and" + finite clause --- not applicable
 --- No clause-level "and" word (209, 230); parataxis used instead (202)
 --- AdvCls formed either paratactically or via subordinate clauses (including RCls) (242-46)
- 15) VN/Inf instead of finite main-clause form --- NO
 --- Both parataxis (195ff.) and serialization (210ff.) use finite verbs
 --- Uses of Inf include no mention of clause chaining (213-14)
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- YES
 --- Word won means both "father of" and "owner of" (158), construed as inalienable possession in both; distinct word "father" pàpò, with no second meaning given (331)

Maasai (East Africa; Nilo-Hamitic = Eastern Nilotic [Eastern Sudanic])
 Tucker & Mpaayei 1955

- 0) Basic categorial information
 --- Noun: gender (masc, fem, place) and number are marked prefixally (15); two cases, Nom and Acc, distinguished tonally (175); Acc used for citation form, verb object, predicate nominal, possessor (175, 213)
 --- Verb: Subj and (usually) Obj person/number marked prefixally (conflated Subj/Obj marker) (71ff.); 3-sg/pl Obj = Ø, 1/2-pl Obj expressed as independent pronouns; gender not marked
 --- Article: gender prefix acts as article; somewhat definite, but translatable either as "a" or "the" (3, 46)
- 1) Conjugated adpositions --- NO
 --- Grammar presents 2 Preps: o "and", te "with/at/for" (41-42); both take normal (independent) personal pronouns (15); 1st and 2nd-person pronouns have special tone pattern after te (216), but no sign of any clitic-like behavior or fusion
 --- Nouns used as Preps (e.g. "top" for "over") govern their objects in Acc (in its capacity as case of Possessor) (216); pronoun objects should thus be the possessive pronoun series (20), again independent words
- 2) Word order
 --- VSO (7); OV possible for emphasis (8, ex and comment)
 --- Preposition (42)
 --- N-Adj (12); a few cases of optional Adj-N (lexically conditioned)
 --- N-Gen (20, 38)
 --- N-RC1 (23)
- 3) Relative clause linker --- ZERO

--- RCl-initial verb begins with special relative prefixes (bound to verb)

4) Relativization strategy/ies --- GAPPING

--- RCl formed by either replacing or preceding the Subj/Obj prefix complex (see [6] below) with special relative pronominal prefixes (23, 106, 109); relative prefix agrees with Head noun in number and gender (23, 106ff., 223ff.)

--- The coreferential full NP deletes

--- For Obj RCl, no pronominal copy survives even on verb (prefix-complex marks only Subj, cf. forms like "I whom you tie", 110); for Subj RCl, verb prefix-complex continues to mark both actants (with modifications); grammar book does not present phenomenon in these terms, but inferrable from discussion (106-110)

--- Possessor RCl simply deletes coreferent NP (113); no mention of Obj-of-Prep relative

--- For verbs with Aux, relative prefixes attach to Aux (112)

5) Special relative form of verb --- YES (special prefixes)

6) Polypersonal verb --- 2 actants sometimes coded on verb

--- Subject: verb prefixes (indep pronoun may optionally follow verb) (53)

--- Object: description (71ff.) a bit obscure, but apparently: 3-pers Object (sg, pl) = zero; combinations with 1-sg and 2-sg Objects expressed via conflated Subj/Obj prefixes; other Objects expressed via independent pronouns following verb (3-pers Obj may be explicitly stated this way too)

--- Imperative has special suffix for 1-pers Obj (71)

7) Infixing/suffixing alternation --- NO

--- There are some Aux verbs (inflected as finite verbs) followed by infinitive of Main verb (96); here the Aux inflects for both Subj and Obj (via the usual conflated markers), except that Object suffix in Imperative attaches to main verb (98)

--- But these aren't preverbs but true verbs, and are inflected AS verbs (thus in RClS, it is the Aux which takes relative forms, see [4] above)

8) Definite article in genitive embeddings

--- Genitives construction involves separate "of" word (like Bantu): HeadN "of" Gen (37-38); "of"-word concords with both HeadN and Gen ("Gen" marked with Accusative case, 213)

--- In N-Gen construction, both nouns take gender prefix, hence both are "definite"; not stated explicitly, but clear from exx., 38

9) Nonconcord of V with full-NP Subj --- NO

--- No explicit statement, but inferrable: grammar states that independent subject pronouns always optional (53), which implies that their presence does not affect verb marking

--- 3sg = 3pl prefix (segmentally); some tenses distinguish sg/pl tonally; some stative verbs do have special plural forms (23), and exx. show that these do concord with full NP

- 10) Verbal abstract: VN or Inf? --- BOTH
 --- Infinitive in such contexts as "I like to V", "he has gone to V" (65); object certainly not in genitive, probably in Accusative case (but no explicit statement is given)
 --- Gerund (distinct from infinitive) in such contexts as "Herding of cattle is good" (66); exx. show object as taking genitive "of" word
- 11) Predicative particle? --- NO
 --- There is a verb "to be something" (91) used for predicative constructions (Adj or N: "be big", "be a chief"); zero in 3-pers Present; no predicative particle
 --- Predicate nominals in Accusative (!) case (175, 196); occur initially in clause, or after "to be something" (12, 196 for Adj; 91, 175 for Noun)
- 12) Prepositional periphrastic (Prep + VN) --- NO
 --- There are Tenses with Aux (96ff.), but none with Prep is mentioned
- 13) "DO" periphrastic ("DO" + VN) --- no mention, assume NO
 --- "DO" not mentioned with other Aux verbs (96ff.)
- 14) Adverbial clause = "and" + finite clause --- no mention, assume NO
 --- There is a clause-level "and" conjunction (103-4)
- 15) VN/Inf instead of finite main-clause form --- YES (65)
 --- Grammar says that Infinitive "may also correspond to 'and' in narration" (65); ex. (235):
- Ole Mpaa was seized and to-lie-on
 "Ole Mpaa was seized and they lay on him"
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- no data

Mandinka (West Africa; Mande [Niger-Congo])
 Creissels 1983; Creissels et al. 1982

- 0) Basic categorial information
 --- Noun: no noun classes (Welmers 1973:184); no case; suffixed Art and Pl, in order: N-Art-Pl
 --- Verb: isolating, no Subj or Obj affixes; finite verbs take "predicative Aux" (conflating aspect, mood, Neg), with order (27):
 SubjNP Aux ObjNP Verb
 --- Article: suffixed Def article (52ff.), special morphophonemics (56), also tonal manifestations (only loosely described, 56-57, 174); "N+Art" is unmarked form, citation form (52); bare N (without Art) may only occur in restricted contexts, notably negatives (52-53); in [N-Adj]

syntagm, Art suffixed only to Adj (94)

1) Conjugated adpositions --- NO

--- There are true Postps (132ff.), relational nouns acting as Postps (133), and nouns used adverbially that may take Objs (138); no pronominal affixes; exx. (134-37 and passim) show independent pronouns as Obj of Postp

2) Word order

--- S-O-V-Other (27); note highly distinctive order
 --- Postpositions (132ff.); only 2-3 elements analyzable as Preps (139), vs. 8 or 9 true Postps, 5 postposed relational nouns
 --- N-Adj (77-79); category "Adj" dubious, but Adj-like modifiers will follow Noun; see pp. 108-9 for morphological treatment of Adj-like roots in predicative and attributional uses
 --- Gen-N (77-79)
 --- RCl: correlative (190ff.)

3) Relative clause linker --- not applicable (correlative)

4) Relativization strategy/ies --- Correlative

--- Two asyndetic, paratactic clauses, in either order (190ff.); Rel marker men, sometimes has plural form (193)

--- Order "RCl-MainCl" (193): HeadN (definite) occurs within RCl in its usual place, followed by "adjectival" Rel marker men, or else have bare men for headless RCl; MainCl has corresponding anaphoric element; thus:

--- Syntagm: you [boy-Art men] saw, I his father know
 Paraphrasable as: Which boy you saw, I know his father

--- Order "MainCl-RCl" (194): "Head Noun" in MainCl is not formally singled out as such; RCl contains men, usually as pronoun; type:

I the boy know, [you men saw] "I know the boy that you saw"

--- Any syntactic role can be relativized (explicit statement, 192); exx. of Subj, Obj, genitival RCl (190) and of Postp RCl (193, 194)

5) Special relative form of verb --- NO

6) Polypersonal verb --- zero actants coded

7) Infixing/suffixing alternation --- not applicable

8) Definite article in genitive embeddings (80-93)

--- Syntagm: Dept (la) HeadN;
 linker la absent esp. with prototypical inalienable possession, present esp. in cases where Dept has power/control over HeadN (84, 92)

--- When HeadN is a nominalization ("the shooting of the hunters"), presence of la shows Subj nominalization, absence shows Obj nominalization (89-91); conforms to criterion of "power/control"

--- With or without la, both NPs may take Art (79; exx. 81, 84); but cannot have indef Head, def Dept (37)

--- Also can directly combine lexical bases (not full NPs) (78), yielding asyndetic Dept-HeadN compound taking Art as a whole (79, 172)

9) Nonconcord of V with full-NP Subj --- not applicable

10) Verbal abstract: VN or Inf? --- INF

--- All verbs can be nominals (23-25); various explicit ways to form verbal abstracts (all forbidding finite-verbal "predicative Aux"):

(i) VERB-ri: found on transitive verbs deprived of ObjNP, yielding an active verbal abstract (89, 125-26, 167); without this suffix, verb is taken passively whenever only one argument is present (cf. English "open", "break"); -ri form takes Art suffix, hence clearly nominal

(ii) Obj-VERB compound (Obj precedes Verb); presented as genitival embedding of Obj (see [8] above), but clearly this is at level of word-formation, not syntax; form takes Art, hence nominal, whereas its incorporated Obj cannot take Art (36, 167)

(iii) VERB-la: follows modal/aspectual verbs in EQUI contexts, or occurs in periphrasis with be (see [12] below) (38-39, 113, 146-47, 172-75); form cannot take Art, but preceding Obj can; ending -la distinct from Postp la, though historically linked (cf. English "to VERB") (146)

(iv) kà...VERB: kà prefixed to entire clause, verb itself unmarked; functions like VERB-la, but also other uses, notably in clause chaining (see [15] below) (147-49); verb does not take Art, but Obj can

(v) Zero derivation: used when verb occurs after "come", "go" (149)

--- Types (i, ii) apparently preclude Obj (except within the verbal abstract); types (iii, iv) take full-NP Obj with normal rection (144-45), hence "Infinitive"; type (v) is found both in the environments of (i, ii) (167) and of (iii, iv) (144-45), behaves in both ways

11) Predicative particle? --- YES

--- Identificational copula mu distinct from situational copula be (161ff.) (176ff., cf. 28ff.); presented as nonverbal but strongly verb-like (30ff.)

--- Identificational syntagm ordinarily requires "focus particle" le after noun conveying new info (33); post-copular Noun (if any) is followed by Postp ti "as, in the capacity of" (135), thus (33-34, 177-79):

N1	<u>le</u>	Cop		"It is N1"
N1	<u>le</u>	Cop	N2 <u>ti</u>	"N2 is N1"
N2 Cop	N1 <u>le</u> <u>ti</u>			(same)

--- Focus particle seems the best identifier of "Predicate" here, can also occur optionally with verbal predicates (33, 176); ti occurs with verbs of change-of-state, e.g. (135, cf. 34)

The man was crowned king-ti;
copular construction seen as same pattern; verb ke "do, happen" may replace Cop (109-10, 176-77), same use of ti

--- Syntagm exactly right, except that ti occurs on final Noun regardless of whether it's "Subj" or "Pred" (new info); perhaps conflicting notions of "predicate" here

--- "PredAdj" apparently behaves either as stative verb (107ff.) or as PredNoun (95-96)

12) Postpositional periphrastic (VN + Postp) --- YES

--- Two types (and a similar third), all using situational Cop BE:

- (a) Subj Cop "VN"-Art la
- (b) Subj Cop (Obj-Art) Verb kan
- (c) Subj Cop (Obj-Art) Verb-la (no Postp, but note -la)

(a) Here "VN" can be either type (i, ii, v) of [10] above; Postp la basically instrumental or purposive, but many nuances (133); Progressive semantics (36-37, 166-68)

(b) "Verb" is ordinary verb base, but behaves nominally (171); kan is Postp "on"; Progressive semantics (37, 170-72)

(c) "Verb-la" is type (iii) of [10] above; Prospective/Future semantics (38-39, 172-75)

13) "DO" periphrastic (VN + "DO") --- YES

--- Construction involving "VN" of transitive verbs (= ri-form, type

(i) of [10] above) plus verb ke "do":

Subj VERB-ri-Art ke

--- Essentially an antipassive; means "Subj does VERB-ing", e.g. "he eats" (Obj unspecified), whereas straightforward verbal syntagm (object-less) would mean "he/it is eaten" (cf. [10] above)

--- Sometimes find bare verb plus ke, e.g. "do work" (54, 60, 131); recall that all verbs can be used nominally (zero-derived)

14) Adverbial clause = "and" + finite clause --- not applicable

--- No mention of clause-level "and" word; none noted in exx. passim

15) VN/Inf instead of finite main-clause form --- YES

--- In chained clauses having same Subj and Aux, verbs in non-initial clauses may convert to kà-Inf (type (iv) under [10] above), with no Subj or Aux (148); clearly nonfinite due to absence of Aux

16) Word-initial change --- no mention, assume NO

--- However, prominent feature of SW Mande (Kpelle, Mende, Loma) (Welmers 1973:130ff.)

17) Extended use of kin terms ("Kin of Noun") --- NO

--- No exx. given in dictionary (Creissels et al. 1982); some kin terms given in grammar (1983:86ff.), again no exx.

Mangarayi (Australia [Arnhem Land]; non-Pama-Nyungan)
Merlan 1982

0) Basic categorial information

--- Noun (56ff.): inflects for noun class (m/f/n), number (sg/du/(trial)/pl), case, conflated in a single paradigm (56) involving both prefixes and suffixes (paradigms 57, 89); expression of plurality often optional (86-87), especially with inanimates

--- Erg-Abs patterning for neut-sg, otherwise Nom-Acc (56, 87)

--- Nouns take possessive suffixes (106): N - Poss - Case

--- Verb: 2 actants obligatorily coded (prefixally) on transitive V (24, 157); allocation of Subj vs. Obj markers to slots not straightforward (161, paradigm 160); Subj and Obj markers pattern Nom-Acc (xiv)

--- Article: anaphor gi-nara (43-45, 93), built on neuter demonstrative pronoun nara (110) (gi- seldom occurs except in this word, 43); anaphor almost always cooccurs with a following noun ([Anaph+N]) whose referent already "belongs" to the discourse, hence specific and identifiable (93) (though the N itself may not have been mentioned, 44); unclear how common this anaphor is (44); plausible candidate for definite article, but will not reanalyze as such

--- Grammar also mentions "specific determiner" -wa (171-72), glossed "Article" but semantically inappropriate for def-article; nothing identified as "indefinite article" (95-96)

1) Conjugated adpositions --- not applicable

--- No category "adposition" (26-27, 166)

2) Word order --- very free

--- Clause-level: apparently free; weak preference for OVS, but very rare to have two full-NP actants at all (25-26)

--- No category "Adposition" (26-27); adpositional notions can be expressed by adverbs accompanied (usually followed) by noun in semantically appropriate "local" case (27)

--- Adj-N or N-Adj, or discontinuous (29, 51); order weakly dependent on lexical choice of Adj (29)

--- Gen-N or N-Gen (30)

--- N-RC1, quite strictly (15, 17, 18) (insofar as category "RC1" is well-established, 12ff.)

3) Relative clause linker --- ZERO

4) Relativization strategy/ies

--- A single "generalized subordinate clause" type (12ff.) serves for Noun-clause, adverbial clause, RC1 (13); characterized by quasi-paratactic, "marginal integration into the main clause" (12); marked only by prefix wa-/ya- (13, 143-44)

--- Interpretation as RC1 is strongly favored if have sequence N-SubordCl (17); relativized N must be cross-coded on RC1 Verb (i.e., must be Subj or Obj in SubordCl) (15); coreferential noun never (?) recurs as full-NP within RC1 (thus "gapping"), but full-NPs are seldom mentioned explicitly even in nonsubordinate contexts

--- Headless relatives common (18-19); also have semi-productive Agent-nominalizer (174-75), and use of Habitual verbs as "attributive actor relatives" (20)

5) Special relative form of verb --- not exactly

--- wa-/ya- an all-purpose subordinator (13)

6) Polypersonal verb --- 2 actants coded prefixally

7) Infixing/suffixing alternation --- NONE

--- Only 36 simplex verbs (131); most verbs involve an inflecting Aux, and follow patterns (i, ii) (123ff.):

(i)	Ptcl + Aux:	Ptcl	Pfx-PronAffx-	Aux	- (Deriv)	- Tns/Asp
(ii)	Compound:	Ø	"	"	-Cmpd+Aux-	"
(iii)	Simplex:	Ø	"	"	- Root	"

--- The elements termed "Ptcl" and "Cmpd", lexically very numerous, convey the real semantics of the verb; but they are not themselves forms of finite verbs, and seldom can occur outside these constructions ("Ptcl" counts as separate word-class) (124, 131)

--- Aux chiefly "a prop for inflectional material" (123); always immediately precedes finite-verb complex (124)

--- Conclude: Aux is not analogous to a Preverb (Preverb would be added to a host which was itself a verb); and even if it were, position of person markers is fixed

8) Definite article in genitive embeddings --- not applicable

--- Normal genitive type (66-67, 106):

Possessor-gen	Head-poss	(either order)
man's	camp-his	"the man's camp"

--- If gi-nara were taken as an Art, note that nouns so marked "are almost always in major syntactic functions" (= Subj, Obj) (44); hence Possessor noun unlikely to take gi-nara

--- Found no exx. of possessive with gi-nara (29-30, 66-67)

9) Nonconcord of V with full-NP Subj

--- V doesn't always show number concord with nonsingular nouns: non-concord commoner with neuters, usually full concord with animates (86, 91)

--- Position of N apparently irrelevant

10) Verbal abstract: VN or Inf? --- VN

--- "Purpose complement" ("in order to V") is non-finite form of subordination (9-11); contrast with finite "generalized subordinate clauses" (cf. [4] above)

--- Purpose-Comp formed by adding Dative/Purposive ending to gerundial form of V; gerundial formed by several productive nominalization processes (172ff.)

--- Obj of Purpose-Comp verb will occur in Genitive (if pronoun) or Dative/Genitive (if full-NP) (11); usually straightforward EQUI syntax

11) Predicative particle? --- NO

--- Construction may (23, 69) or may not (62-63) have overt Copula

--- If overt copula, PredNom is nominative or caseless (23, 69); if no copula, PredNom itself inflects as intransitive verb, or takes only number markers (62-63); in neither case is any particle involved

12) Adpositional periphrastic (Adp + VN) --- not applicable (no Adpos)

13) "DO" periphrastic (VN + "DO") --- in a sense

--- Commonest Aux verb is ma- "do, say" (123, 129); but (see [7] above) the lexical complement (Ptcl or Cmpd) of Aux is not itself deverbal

--- English verbs can be borrowed as Ptcl+Aux, e.g. wurg ma- "work" (lit. "do work"); borrowings usually coupled specifically with ma- as Aux (129)

14) Adverbial clause = "and" + finite clause --- not applicable

--- No plain conjunction "and" (31); adverb clauses built with "generalized subordinate form", with no conjunction (20ff.)

15) VN/Inf instead of finite main-clause form --- weakly

--- Clause-chained verbs are normally fully inflected (164); but can chain "Ptcl+Aux" verbs by omitting Aux and giving just the Ptcl (165); this is nonfinite, though the nonfinite form (the Ptcl) is not a VN (see [7])

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- seemingly not

--- No examples readily apparent in "Lexicon" section (222-35)

Maricopa (Arizona; Yuman [Hokan])

Gordon 1986; also Langdon 1977

[My thanks to Pam Munro for answers to several queries]

0) Basic categorial information

--- Noun: no gender; suffixal case, with Acc and Gen zero-marked (36), oblique case-markers optional (48); an NP modified with {Gen, Adj, RCl} is only case-marked once, at end of NP (250, 255, 261); some nouns (animates only) have distinct plural forms, usually formed by stem Ablaut, sometimes with affixes (29); pronominal possessor (inalienable possession) coded prefixally (30ff.), with plural suffix -sh on Noun denoting plurality of possessor (34); demonstrative suffixes (53); slot sequence:

Poss-Noun-Dem-Case

--- Verb: prefixal person coding of Subj of intrans (16), partially conflated Subj/Obj of trans verb (18ff.) (table 21); codes IObj rather than DObj if 3 participants (42); verbs have distinct plural-Subj and dual-Subj forms (formed by Ablaut, affixes), and sometimes a plural-Obj prefix nyi- (22-23, 90ff.); slot sequence

nyi-Subj/Obj-Verb;

final suffix -k, -m plays a variety of roles, notably switch-reference in subordinate clauses (24); oblique case markers (or sequence "Dem+Case") can detach from NP and procliticize to Verb (50, 158ff.); extensive V+Aux combinations, employing Subj-marking (on both verbs) and Switch reference (183-84)

--- Article: demonstrative suffix -ny glossed 'anaphoric, definite, generic' (53); very frequent in texts, commonly suffixed to nominalized clauses, but not obligatory to convey definiteness (53); will analyze as "hedged article"

1) Conjugated adpositions --- irrelevant (perhaps yes)

--- No category "adposition" mentioned; "locative nouns" exist ("the tree's top", etc.) but not restricted to adverbial function (46-48); these take Obj as directly juxtaposed Genitive; no discussion of pronominal Obj, but would presumably be possessive prefixes (see [8]); will not count these as "Postpositions"

2) Word order

--- SOV; V always final, NPs flexible (15, 74)

--- No category "Adposition"; postposed "locative nouns" (46-48)

--- N-Adj; but "Adj" are verbs, in special type of RC1 (see [4] below) (28, 51)

--- Gen-N (31)

--- N-RC1, or irrelevant (various strategies, see [4] below)

3) Relative clause linker --- NONE

4) Relativization strategy/ies

--- Various strategies (254-65):

(a) Certain intransitives ("adjectives") follow HeadN with no special mark (51, 250)

(b) For intransitive Subj-RC1, can use "switch-reference-marked" RC1s: RC1 verb follows HeadN, marked for switch-ref vis-a-vis the following main-clause verb; HeadN case-marked either for RC1 role (= Subj) or matrix role (250-52)

(c) Correlative strategy (not so named): two separate clauses, in order RC1-MainCl; RC1 verb must take "emphatic perfective" aspect suffix -(k)sha (not a special relative form, 111); no mark in RC1 signalling which NP is conceptual HeadN; HeadN is resumed pronominally in main clause (252-54)

(d) Nominalization (internally headed): RC1-verb assumes nominalized form (see [10] below), Subj within RC1 takes zero case-mark (normal in Nlztn); HeadN remains in situ (though may be fronted within RC1), takes case-marking appropriate to embedded clause; for Subj-RC1, RC1 verb takes special marker kw as Subj prefix (254-65)

--- Schematically:

(a) [HeadN-Ø Verb]-Case

(b) [HeadN-(Case) Verb-SwRef] ... MainVerb

(c) RC1 MainCl (with resumptive pronoun in MainCl)

(d1) [HeadN-Ø ...kw-Verb.Nlz]-Case (Subj RC1)

(d2) [... HeadN-Casel... Verb.Nlz]-Case2 (non-Subj RC1)

--- In (a) and (d), case of HeadN in matrix clause is coded once, at end of RC1

--- No category "Adp", hence no Adp-RC1

--- Genitival RC1s only possible by "Possessor Raising" (258-60), a process which exists only for intransitive verbs and only for possessor of Subj: possessor "raises" to become Subj of clause (not just RC1) (68-71), e.g. (here "black" is a verb):

[you your-hair] it-black ==> you hair you-black "Your hair is black" with genitival RC1 thus recast as Subj RC1

5) Special relative form of verb --- YES, sometimes

--- Nominalized Subj RC1s marked with kw; no special mark otherwise

except nominalization, which has other uses

6) Polypersonal verb --- 2 actants coded

7) Infixing/suffixing alternation --- NO

8) Definite article in genitive embeddings --- hedged

--- Three Gen patterns (31-33), always with DeptN taking zero ending (40):

- | | | | |
|---------------------------------|----|--------------------------------|-----------------|
| (a) DeptN HeadN-case | OR | Pfx-HeadN-case | (inalienable) |
| (b) DeptN <u>ny</u> -HeadN-case | OR | Pfx- <u>ny</u> -HeadN-case | (certain nouns) |
| (c) DeptN HeadN <u>nywish</u> | OR | HeadN Pfx- <u>nywish</u> -case | (other nouns) |

--- Plurality of DeptN may be indicated by suffixing -sh to HeadN (34)

--- No statement of distribution of "article" -ny in Gen; hardly any relevant exx.: no ex. shows two Arts, one ex. with Art only on Dept (35), one ex. with Art only on Head (68); Art may occur on prefixally possessed nouns (68, 71)

--- Art on Head is completely normal in Yuman [PM, p.c.]; Art may also cooccur on Dept, but much more unusual [PM]

9) Nonconcord of V with full-NP Subj

--- Most nouns (and indep personal pronouns) lack special pl forms (29, 58), whereas verbs have distinct forms for plural-Subj and for dual-Subj (22, 90ff.); but may use sg Verb-form with non-singular Subj (21-22), and may refer to a pl entity with sg Noun-form even when a special pl form exists (29); should imply flexibility in overt concord [my own argument; no explicit statement]

--- Plural nouns probably are preferred if they exist, and plural verb-forms probably are preferred when appropriate [PM, p.c.; not sure]

--- Exx. of conjoined Subj showing verb concord (67, 99); one ex. of counted noun as Subj showing verb concord (51)

10) Verbal abstract: VN or Inf? --- analyze as NEITHER

--- Verb may be nominalized by various Nlzd affixes (227ff.); Nlzd clause takes case-mark per its role in matrix clause, case-marks its own full-NP Subj not with usual -sh but with zero (like a possessor), and may negate like a noun, not a verb (230); all other NPs in Nlzd clause case-marked as usual (per exx. passim)

--- However, nominalized verbs retain all Subj/Obj markers of finite verbs; possible ambiguity, since 3pers-Subj is zero anyway, and intransitive Subj markers = possessive markers; but exx. with conflated Subj/Obj marker (e.g. 230, "your hitting me") show unambiguous verbal coding

--- This fits criteria for calling nominalization and SwRef "finite forms"; such an analysis differs strongly from Yumanists' usage [PM, p.c.]

11) Predicative particle? --- NO

--- Use copular verb, no particle; curious case-usage, with nominative case -sh on Pred (not on Subj) (38-41, 171):

Subj-Ø Pred-sh (Cop)

- 12) Postpositional periphrastic (VN + Postp) --- not applicable
- 13) "DO" periphrastic (VN + "DO") --- in a sense
 --- "Expressive constructions" (178-80): X + "say"
 with X uninflected; causative version is: X + "do"
 --- Langdon presents Pan-Yuman discussion of phenomenon (not mentioning Maricopa): X sometimes may be a form of an ordinary verb (1977:1; a few exx. passim), but clearly not systematically deverbal; no mention in Gordon
 --- Also have "do" as Aux, but preceding verb is finite (181ff.)
- 14) Adverbial clause = "and" + finite clause --- not applicable
 --- Section "Clausal Conjunction" mentions no clause-level "and" word (284ff.)
 --- Adv Clauses by clause-chaining (277) or verbal affixes (266ff., 148-50)
- 15) VN/Inf instead of finite main-clause form --- not applicable
 --- No nonfinites; extensive clause-chaining with SwitchRef, but all verbs are equally finite (115ff.)
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- apparently NO [PM, p.c.]

Lake Miwok (California; Utian [Penutian])

Callaghan 1963; Callaghan 1965; see also "bibliographical note" below
 [My thanks to Catherine Callaghan for advice and clarification]

0) Basic categorial information

--- Noun: takes possessive prefixes (table, 75), optional dual and plural suffixes (rare for nonhumans) (77-78), case suffixes (list 87; paradigms 122-23)

--- Verb: grammar presents language as having subject prefixes (59; table 125), but probably better taken as clitics [CC, p.c.], and not part of inflection (see [6]); Sierra Miwok shows considerably more head-marking [CC, p.c.]

--- Article: no mention; pronominal element ne- "this" (58, 1965:100), which may be either independent pronoun or first element in nominal compound, is evolving toward articlehood [CC, p.c.] (exx. 79, 1965:28), becomes still more Article-like in Bodega Miwok; hardly obligatory in definite contexts, though not infrequent in 3 texts I checked (see below), but often introducing new information; will not reanalyze as "hedged article"

--- Bibliographical note: information in the grammar was augmented by consulting 3 texts from volumes of Native American Texts Series (IJAL), viz. Northern California Texts (1977:10-16), Coyote Stories (1978:62-86), and Coyote Stories II (1980:81-87)

- 1) Conjugated adpositions
 - No category adposition; nouns may be used instead (1977:16 note 10), but these can also serve adverbially, don't make a well-profiled category [CC, p.c.]
- 2) Word order
 - SOV unmarked order (3, 256); SV far commoner than VS (92); VO order for emphasis, afterthought, etc. (97, 257)
 - No category adposition
 - Adj-N (66-67); Adjs do count as a distinct part of speech [CC, p.c.]
 - Gen-N (94ff.); no statement, but thus per all exx.
 - N-RC1, though no statement; RC1 essentially headless, but all headed exx. apparently show order Head-RC1 (see [4])
- 3) Relative clause linker --- ZERO
 - Few headed examples anyway; special RC1-marker -hinte occurs clause-finally on verb, hence does not function to link HeadN and RC1; see [5]
- 4) Relativization strategy/ies --- GAPPING
 - Barely described in grammar (254-55); description inferred from exx. (93-94, 101, 107, 108, 112, 113, 118); exx. in the 3 texts show a few puzzling (and unexplained) uses of final case endings on RC1s, clashing with account to be presented here
 - RC1 verb apparently final in clause; takes suffix -hinte, which then (usually?) takes case suffix appropriate in matrix clause
 - Most exx. headless; only 3 headed exx. in grammar (94, 108, 112), showing order N-RC1, with Head and RC1 taking identical case-mark (hence appositional?); dictionary has one ex. where HeadN and (following) RC1 are discontinuous, with main-clause verb intervening (1965:35); in texts, all (?) headed exx. appear to have order N-RC1 (possibility of internal headedness?)
 - Coreferential noun gaps, apparently
 - No exx. of genitive RC1s in grammar or texts; status uncertain [CC, p.c.]
- 5) Special relative form of verb --- YES
 - Particle -hinte, suffixed to verb, specialized for RC1 function (254)
- 6) Polypersonal verb --- zero actants (depends on analysis)
 - The so-called "verb subject prefixes" are better taken as clitics [CC, p.c.], not part of verbal inflection at all, because they:
 - (a) May appear (certain allomorphs) enclitic to previous word (126)
 - (b) May be preposed to a phrase (with non-initial verb) (59)
 - (c) Usually don't cooccur with a 3-person noun Subj [CC, p.c.]
 - (d) Can be omitted, esp. in narrative, and esp. 3-sg [CC, p.c.]
- 7) Infixing/suffixing alternation --- not applicable
- 8) Definite article in genitive embeddings --- not applicable

--- No article; Gen syntagm is: Possessor-gen Head
(per exx. 94ff.); no genitive exx. with ne- in grammar (and I noted none in texts)

9) Nonconcord of V with full-NP Subj --- not applicable

--- Subj clitics usually do not cooccur with full-NP subj [CC, p.c.], so agreement of full-NP Subj with verb not an issue; clitics do cooccur with an independent non-3-person pronoun Subj [CC, p.c.]; but clitics aren't analyzed as part of verb inflection

10) Verbal abstract: VN or Inf? --- not applicable

--- Some contexts apparently favor omission of subject clitic, and verb may itself take case-markers (implying nominalization of clause); but since subject clitic not analyzed as part of verbal inflection even for "finite" verbs, have no morphological finite/nonfinite distinction

--- All instances where subject is omitted show normal (non-genitive) rection of Obj: "He wants to catch me"; "That's our job, killing people" (255)

11) Predicative particle? --- NO

--- Equational sentences are verbless [CC, p.c.]; no particle; NomPred apparently occurs in Appositive case (88-89), apparently in the order
Subj - NomPred (per exx.)

12) Prepositional periphrastic (Prep + VN) --- not applicable

--- No category "adposition"

--- Durativity can be rendered by Instr (a "local" case) on verb (116):

VERB-Instr + MainVerb ;

but Instr mark applies to clause as a whole; embedded verb can indicate subject as usual (per ex. 116), hence "finite"

13) "DO" periphrastic (VN + "DO") --- apparently yes

--- Dictionary (1965:79) has example with locok "do":

ritual sing-Instr they-do = "They sing the ritual (song)"

--- Periphrases with hinte "do" common [CC, p.c.], though no exx. in dictionary

--- Note also the following syntagm (1965:156) (cf. [12] above):

VERB-Instr + weeTa "go" = "to be doing something"

14) Adverbial clause = "and" + finite clause --- YES

--- Particle miTi: if clause-initial, presented as an "introductory particle" and glossed "and" (245, 252); if clause-final, glossed "when, after" (250, 255-56)

--- Possible ambiguity [my own suggestion] in frame:

John came miTi I invited him in.

15) VN/Inf instead of finite main-clause form --- not applicable

--- No nonfinite form; Subj clitics omissible in narrative [CC, p.c.], and 3-sg clitic frequently omitted in example sentences in grammar [my own observation]; but clitics not part of verb inflection

16) Word-initial change --- NO [CC, p.c.]

17) Extended use of kin terms ("Kin of Noun") --- NO [CC, p.c.]
 --- No exx. in dictionary s.v. "father, mother, child" (Callaghan 1965)

--- But in Bodega Miwok (closely related), "father" is used regularly as personal agentive [CC, p.c.]; exx. from Callaghan's Bodega Miwok Dictionary, UCPL 60 (1970), p. 88, s.v. "father" (?ap(p)i): "father of running" = racer, "father of chopping wood" = woodchopper, "father of sucking" = suck doctor, "father of stealing" = thief, "father of killing" = murderer

Mixtec (Ayutla) (Mesoamerica; Otomanguan)
 Hills 1990

(with notes on other Mixtec dialects from Bradley & Hollenbach 1988, 1990)

0) Basic categorial information

--- Noun: no cases; no gender; number seems not to be marked
 --- Verb: no affixal person/number marking; verbs and other Heads may take as arguments either nouns or pronouns (free or clitic) (209-10); unclear if clitic should count as argument-marking on the verb, but probably not; patterning of free and clitic pronouns as follows:

(a) 1/2-person pronoun: free and enclitic forms both exist; pronoun serving as verb Obj must be free form, all other roles may be either free or clitic

(b) 3-person pronoun: clitic forms only (usually enclitic); can serve any syntactic function except 3-f-sg Obj (here a special free form is used); nothing said about cooccurrence of Subj and Obj clitics

--- Article: none (though "one" is usable as indef article, 120)

1) Conjugated adpositions --- in a sense

--- Only 5 prepositions (214-15); most prepositional functions are fulfilled by locative nouns, analyzed here as "hedged prepositions"; both Preps and locative nouns may take either free or clitic pronominal objects (see [0] above)

--- Other Mixtec dialects tend to show some explicit fusion between (some of the) clitics and Prep, arguing more strongly for "conjugated Prep"

2) Word order

--- VSO (27)

--- Prepositions (214)

--- Adjectives: category does not exist (use stative verbs) (22, 127)

--- N-Gen (133)

--- N-RC1 (122)

3) Relative clause linker --- ZERO (123)

--- Some dialects have non-zero linker: a complementizer, or a form homonymous (identical?) to a pronoun

4) Relativization strategy/ies --- basically GAPPING (123)

--- RCl identical to main clause except for gap of coreferential NP (123)

--- For Subj RCl, two types: (a) N RCl; and (b) N Pron RCl, with Pron coreferential to N. Type (b) is analyzed as "appositional" to N (with Pron considered to be true head of RCl; Pron-RCl appositionally restates N). "Appositional" RCl (b) are for restrictive RCl in Ayutla (122), for non-restrictive in Jamiltepec Mixtec (I:68); just the inverse holds in the two dialects for type (a). The appositional type looks like a resumptive SUBJECT pronoun (?), but not described as such.

--- Can relativize on Subj, Obj, Time/Place, or Obj-of-Prep (124-25); if Obj-of-Prep, strand Prep (or locative noun) in situ in RCl, like English "dangling Prep"; no mention of possibility of relativizing on Possessor

--- Coatzacoapan Mixtec does have resumptive pronoun on stranded Prep (II:349); but (in non-relative clauses) Prep can be "incorporated" into V nucleus, i.e. fronted to just after V (II:281, 312), in which case the corresponding prepositional RCl has no resumptive Pron (II:350)

--- Ocotepec Mixtec, in prepositional RCl built on Prep "with", fronts the (bare) Prep to just after Subject (I:214); otherwise "dangling Prep" in situ

5) Special relative form of verb --- NO

6) Polypersonal verb --- apparently no true arguments coded on verb

--- Subj, Obj can be noun or pronoun (exx. 12ff.); Subj pronoun can be free or clitic (with restrictions, see [0] above), Obj pronoun only free (209-10); no statement as to whether clitic pronoun can double a full-NP argument, but all (?) exx. passim seem to indicate not; pronouns are presented as an optional realization of NP (117), not as part of VP; in some dialects, some clitics can fuse with a preceding stem (e.g. a verb)

7) Infixing/suffixing alternation --- not applicable

8) Definite article in genitive embeddings --- not applicable

--- Genitives by simple juxtaposition (133); no article

9) Nonconcord of V with full-NP Subj --- not applicable

10) Verbal abstract: VN or Inf? --- NEITHER

--- Finite verbs in prototypical infinitive contexts (e.g. "want", 34-35); no mention of any abstract verbal

--- "Most languages in this family [Mixtecan] lack sentences that are untensed, such as infinitival sentences" (Bradley & Hollenbach 1988:6)

11) Predicative particle? --- NO

--- Adjective predicates are stative verbs (22)

--- Noun-predicate clauses formed with Equative verb: EQ.VB Subj Pred (21); order differs in other dialects

--- Verbless clauses do exist in Coatzacoapan Mixtec (II:283), but not in most dialects

12) Prepositional periphrastic (Prep + VN) --- not applicable

13) "DO" periphrastic ("DO" + VN) --- not applicable

- 14) Adverbial clause = "and" + finite clause --- weakly (semantics)
 --- te "and" introduces purpose clauses; formally identical to coordinate clauses (239)
 --- In Jamiltepec Mixtec, "and" for result clauses (I:131)
- 15) VN/Inf instead of finite main-clause form --- not applicable
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- apparently NO
 --- No exx. in 4 brief dictionaries of different Mixtec dialects supplied by Leanne Hinton [p.c.]

Nkore-Kiga (East Africa; Bantu [Niger-Congo])
 Taylor 1985

- 0) Basic categorial information
 --- Noun: noun-class system (prefixal) conflating class and number (121ff., table on 124); class prefix normally preceded by a preprefix (termed "initial vowel", 88), occurring except under specified conditions (88-89), and not itself a marker of definiteness (125); no cases
 --- Verb: Subj and two Objs coded prefixally, with DObj coding = IObj coding (order is Subj-DObj-IObj, 171); full-NP Obj will be coindexed on verb (Obj affix) iff full-NP Obj precedes the V (170-71); table of affixes (130)
 --- Article: no article as such (53); but initial vowel, when present on a postnominal adjective, codes definiteness on the NP (= [N Adj]) as a whole (125-26); ditto for initial vowel in adnominal RC1 (i.e., initial vowel of Rel-particle or of Rel-prefix) (21-22, ex. 176); will analyze as "hedged article"
- 1) Conjugated adpositions --- YES
 --- "Strictly speaking, the preposition does not exist as a separate class" (86); but four particles ("on, in, with, like") act as Preps (86), and these do combine with personal pronouns (182); will not count "of" as a Prep
 --- Also have "compound Preps" (86, 181), of the type
 Prep + Noun + of e.g.: "on ground of" (= under);
 here "of" combines with personal pronoun
- 2) Word order
 --- SVO (45); Obj can precede verb (90)
 --- Prepositions (51, 86, 180-82)
 --- N-Adj (54, 85, 174-76)
 --- N-Gen (100)
 --- N-RC1 (23)
- 3) Relative clause linker
 --- Rel particle coding class/number (Non-subj RC1); zero (Subj RC1)

- 4) Relativization strategy/ies --- GAPPING and COPYING
 - All RCls: change regular Tense affix to a special subordinative-tense affix (167-69), termed "participial" but fully finite (25), used in almost all subordinate clauses (Noun clause, Rel clause, Adv clause); not all tenses have a distinct participial form
 - Subject RCl: gap noun; change verb's Subj Class-prefix to special relative prefix, usually consisting of Class-prefix preceded by initial vowel (141-42; table, 142)
 - Non-subject RCl: precede RCl with a relative particle, agreeing in Class with HeadNoun (23, 142) and always 3-person (143); verb unchanged (except for assuming "participial" form)
 - Obj RCl: gap noun, delete Obj affix on verb (142)
 - Possessive RCl: change coref possessor-NP to pronominal copy in situ (24-25); not clear what happens if coref possessor-NP occurs as part of Subj, e.g. "the man [whose child just died]"
 - Prep RCl (24, 143): either delete Prep and its object entirely; or delete object and recast Prep as enclitic on Verb (51, 190-91); or (for compound Preps) create pronoun copy of object in situ (143)
 - All RCls: relative prefix and relative particle may either include or omit initial vowel, thereby coding definite vs. indefinite RCl (21-22, ex. 176)
- 5) Special relative form of verb --- YES (Subj RCl only)
 - Subject RCl has special forms of prefix; many tenses employ a special "participial" form (which, however, is used for all sorts of subordination)
- 6) Polypersonal verb --- 3 actants coded prefixally
 - Subj, Dobj, IObj (exx. 15, 43, 171)
- 7) Infixing/suffixing alternation --- NO
 - Position of Obj infixes unchangeable
- 8) Definite article in genitive embeddings --- hedged
 - Possessive syntagm (100) is: [Cl-]Head Cl-a Possessor, where "Cl" means "noun-class marker" and the "of"-word -a agrees in class with Head
 - Exx. (100) show Head-N with initial vowel intact; but found no exx. involving Adj, hence no exx. relevant to definiteness
- 9) Nonconcord of V with full-NP Subj --- NO
 - Full-NP Subj always marked on V, always concords (170); note that full-NP Obj is coindexed on V only if Obj precedes V (171)
- 10) Verbal abstract: VN or Inf? --- INFINITIVE
 - The only nonfinite form is infinitive (169); exx. (169) show Inf with normal (nongentive) marking of full-NP Obj; Inf with pronominal Obj codes Obj in usual slot in the verb (29, 173)
- 11) Predicative particle? --- NO
 - Syntagm (37): Subj Copula Complement; copula obligatory, no particle

- 12) Prepositional periphrastic (Prep + VN) --- NO
 --- There are compound tenses built on "be" (155ff.), but no Prep is involved and both verbs are finite
- 13) "DO" periphrastic ("DO" + VN) --- no mention, assume NO
- 14) Adverbial clause = "and" + finite clause --- no mention, assume NO
 --- Discussion of coordination (55-56) mentions various "and" words, but no adverbial use; discussion of Adv Clauses stipulates presence of an "adverbial conjunction" (26)
- 15) VN/Inf instead of finite main-clause form --- no mention, assume NO
 --- Nonfinite forms "are normally found in subordinate clauses and seldom form independent clauses" (96)
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- seemingly not
 --- Kin terms discussed (100-101, 223ff.), no figurative use mentioned; no examples readily apparent in "Lexicon" section (226ff.), including a long list of figuratively named plant-names (236-47)

Dongolese Nubian (Northern Sudan; Eastern Sudanic)
 Armbruster 1960; Armbruster 1965

- 0) Basic categorial information
 --- Noun: no gender (130); plural and case marked suffixally (132, 158-59):
 Noun - Pl - Case;
 case marked only on final element in NP (327)
 --- Verb: Subj marked suffixally, per pattern: 1sg, 2/3sg, 1/2pl, 3pl (195ff.); Indep personal pronoun optional (346); verbs have special "plural-Object" forms, formed by suffix -Ir (203, 375), but Obj not marked on verb (no explicit statement; see exx. 375); verbs presented as having a "Subjunctive" form for use in subordinate clauses (195, 370), but appears formally identical to Indicative except for addition of case endings (per morphological description, 196-97)
 --- Article: no definite article; indef article is enclitic form of numeral "one" (165, 323, 338)
- 1) Conjugated adpositions --- apparently YES
 --- Postpositions govern various cases (299ff.); pronominal Objects show some phonological assimilation to initial of Postp (173), part of more general process of assimilation (49ff.), but pronominal objects seem to show special assimilations (see esp. 62)
- 2) Word order
 --- SOV (317-18), apparently quite rigid
 --- Postpositions (299)

- N-Adj (320)
- Gen-N (319)
- N-RC1 (427)

3) Relative clause linker --- ZERO (427)

4) Relativization strategy/ies

--- The material presented under "relative clauses" (427-29) mentions only DObj RCls; syntagm is:

HeadN [RC1]-(Pl)-Case ,

with Obj gapped; verb is characterized as "Subjunctive" (see [0] above); "Pl" optionally present if HeadN is plural; "Case" indicates case of HeadN in matrix clause, suffixed to entire RC1 in conformity to general case-placement rule ([0] above)

--- For equivalent of Subj RCls, use "Participle" (198), termed a "nomen agentis" but usually equivalent to English RC1 (371); participle postnominal, like any Adj; past and present participles exist, both formed with suffix -(I)l (198, 202)

--- For equivalent of temporal and locative RCls, use Noun-clause in genitival embedding to a following loc/temp HeadN (427):

[NounClause]-gen HeadN ,

with HeadN gapped; not presented as "relative clause"

--- No information on Postpositional or Genitival RCls; exx. (438) where coordinate clauses render English RCls, including one genitival ex.:

There was a man, his dog was with him
= "There was a man whose dog was with him"

5) Special relative form of verb --- YES, sometimes

--- "Subjunctive" not a formally distinct form

--- "Participle" devoted chiefly (entirely?) to Subj RC1 function (371)

6) Polypersonal verb --- one actant (Subj)

7) Infixing/suffixing alternation --- not applicable

8) Definite article in genitive embeddings --- not applicable

--- No definite article; syntagm is: Dept-gen HeadN ;
nothing may intervene between Dept and HeadN (319); see also 162-63, 322-23

9) Nonconcord of V with full-NP Subj

--- Full concord is the rule (367)

--- Collectives, counted nouns, conjoined nouns may sometimes take sing or plural concord (367-68)

10) Verbal abstract: VN or Inf? --- BOTH

--- "Infinitive" presented as stereotyped 3pl form of "Subjunctive" (198) or (occasionally) of Conditional (374), followed by case ending showing role of "Inf" in matrix clause; construction just like that of finite Noun Clause, often liable to confusion with it (422):

[Inf / NounClause] - Case (374, 421-22)

--- Inf characterized as inherently lacking its own full-NP subject

(374); indeed, exx. (EQUI etc.) appear passim (e.g. 374) where a 3pl interpretation is semantically anomalous, arguing that "Inf" has gained an autonomous existence despite formal identity with finite NounCl; for NounCl see 421-27

--- Inf governs its Obj just like finite verb (374)

--- Also have a true verbal abstract: a Verbal Noun in -ar or -id, very productive (141, 145), takes Obj in genitive (325, 374); has full case inflection (212)

11) Predicative particle? --- YES

--- Invariant "Predicative suffix" -n (132, 296-97, 401ff.), functions in lieu of copular verb; not itself a verb but a kind of postpositional element (296); suffix used for all pers/numb except 1sg (which instead uses a true 1sg copular verb form Eri, 296); -n must attach to Pred, which may be Noun or Noun-equivalent or Adv (401); syntagm is:

Subj Pred-n

--- This same element alleged (132) to recur in 2/3-sg and 3pl verb endings (= -n, -ran, respectively); but surely might easily be coincidence; apart from this, finite verb cannot take PredParticle (401)

--- Other predicative particles exist too (definite predication, interrogative predication) (296-97, 401ff.)

--- There are various true "be" verbs (379-81), especially E "say, be"

--- Various types of Adverbial Pred can take zero Cop (408, 413)

12) Postpositional periphrastic (VN + Postp) --- no mention, assume NO

--- Progressive rendered as compound verb-stem based on ag- "squat, be present": ag-Verb = "be engaged in Verbing" (266, cf. 380)

--- Discussion of syntax of Postp mentions syntagms N+Postp, Clause+Postp, but fails to mention Inf/VN + Postp (299)

13) "DO" periphrastic (VN + "DO") --- in a sense

--- No periphrastic with "do" mentioned anywhere (see esp. 380-82; also discussion and exx. of "complex verbs", 257ff., 265)

--- Numerous composite verbs of type: X + E; here E = "say, be", and X is invariant (192), the two fusing into a single composite stem; X often onomatopoeic but need not be, need not involve sounds at all; X cannot occur independently, is "certainly not a verb" and (apparently) is not deverbal; type is compared to Amharic, called "one of the most typically Cushitic features of Nubian" (192; cf. also 30ff.)

--- This type near-universal with borrowed Arabic verbs (246)

14) Adverbial clause = "and" + finite clause --- trace?

--- Clause coordination done either asyndetically, or by "and"-Postp added (apparently) to Subj of second clause (437-38; cf. 306-7, 336ff. on syntagm [Noun-"and"], which grammar terms "noun-concretion")

--- Explicit statement that one of the coordinated clauses may correspond to an English subordinate clause (438); but the single ex. given is asyndetic

15) VN/Inf instead of finite main-clause form --- no mention, assume NO

--- No mention under uses of VN and Inf (see [10] above), or under subtypes of one-word sentence (317)

16) Word-initial change --- no mention, assume NO

- 17) Extended use of kin terms ("Kin of Noun") --- NO
 --- Dictionary (Armbruster 1965) apparently gives no exx.

Persian (modern) (Iran; Iranian [Indo-European])

Lambton 1974; also Amin-Madani & Lutz 1972

[My thanks to Mohammad Dabir-Moghaddam for answers to several queries]

0) Basic categorial information

--- Noun: no gender (4); no case (4); plural suffix -hā or (human) -ān (8); in HeadN-Adj or HeadN-Gen constructions, HeadN takes "ezafe" suffix -e as linker (9); suffixal articles (see below); pronominal possessor may be coded as pronominal suffixes (29), with order:

N - PronSuff - Art (29)

--- Verb: Subj coded suffixally (11ff.); pronominal suffix may encode Obj (30), also (in colloquial language) 3sg Subj of intrans. (173); numerous compound tenses with Aux (17-18); present tense of "BE" can be enclitic (12)

--- Article: suffixal indefinite Art -i (3, 125-28); suffixal definite Art -rā, but only for Obj (usually DObj) (4, 130-32); Indef and Def article may cooccur, to express "a certain" (132); article appears once only, at end of N-Adj or N-Gen syntagm (9, 20), but only on HeadN in RCls (no explicit statement)

1) Conjugated adpositions --- YES

--- Obj may appear either as Pron Suffix or as Indep Pron (30); suffix characteristic of spoken language [MDM, p.c.]

2) Word order

--- SOV (5)

--- Prepositions (110ff.)

--- N-Adj (19), linked by ezafe

--- N-Gen (9), linked by ezafe

--- N-RC1 (75ff.)

3) Relative clause linker --- INVARIANT ke (75)

4) Relativization strategy/ies --- GAPPING, Copying (Prep, Gen RC1)

--- Syntagm: HeadN ke RC1 (75), where RC1 is normal finite clause

--- For Subj RC1, gapping; for DObj RC1, gapping or copying; for Prep RC1 (including IObj), copying; for Gen RC1, copying (1972:397-99)

--- If HeadN is definite, usually add -i to HeadN (75), but only if restrictive RC1 (77); this creates def/indef ambiguity (77); further, definite HeadN may optionally take -rā if functioning as DObj in either clause (and as Subj in the other) (76), thus:

HeadN-i-rā ke [RC1]

--- If HeadN+RC1 is predicate of a copular sentence, HeadN and clause are separated by copula (77): e.g.,

This [the same person] is who [RC1] = This is the same person who...

- 5) Special relative form of verb --- NO
- 6) Polypersonal verb --- 2 actants maximum
 - Subj coded suffixally; Obj may be coded by pronominal suffixes or as free forms, with Obj suffix characteristic of spoken language [MDM, p.c.]
 - Note postverbal position of Obj suffix, precluding analysis as a cliticized full-NP (which would be preverbal) [my own argument]
- 7) Infixing/suffixing alternation --- no mention, assume NO
 - In compound verbs, Obj suffix appears not on main verb but on the preceding nonfinite part (93): X-Sfx Verb; but not the right construction
- 8) Definite article in genitive embeddings
 - Def article may only occur on Objects, not on Genitives; syntagm for Objects would be: [HeadN-e Gen]-rā (9), with rā at end of entire NP as usual (and with ezafe as usual)
- 9) Nonconcord of V with full-NP Subj --- for inanimates only
 - Plural Subj takes plural verb if denoting rational beings, singular verb otherwise (13); distinction less rigorous today than in Classical Persian
 - Collectives of rational beings ("army") can take sg or pl verb (133)
- 10) Verbal abstract: VN or Inf? --- VN
 - Persian has a form called "Infinitive"
 - In Modern Persian, Inf is never dependent on another verb; rather, Subjunctive is used instead (1972:491); but Inf can be used as verbal noun (492), functioning as Subj, Obj, PredNom, Adv (492-94); all exx. of "VN" show Obj following Inf and linked to it by ezafe, which precisely matches genitival syntax; see also 1974:143-44 (with more exx.), which characterizes this construction as "Infinitive ... used as nouns"
 - Inf cannot take a preceding Obj marked with Def-Obj marker rā [MDM, p.c.]; but a generic Obj can be "incorporated" into verb (finite or nonfinite) before verbal stem, yielding an Inf which appears to have a preceding Obj [MDM, p.c.]
 - In Classical Persian, can find true infinitival syntax, i.e. with Obj preceding Inf (e.g. subordinated to modal verb) and marked with rā; thus 1972:491-92, 1974:143-44 (with exx.)
 - NB: in cases where "Inf" has its Obj as pronominal suffix, rection is indeterminate: same suffixes serve for pronominal possessor and verb object
 - There are also "Verbal Nouns", e.g. in -eš (1974:96ff.); these are highly nominal forms, clearly derivational and lexical, whose semantics sometimes may correspond to verbal abstract, sometimes to concrete noun [MDM, p.c.]; -eš used today to coin new words [MDM, p.c.]
- 11) Predicative particle? --- NO
 - Syntagm is: Subj PredN/Adj Cop (6, 12); Copula may be either free form or enclitic (12)
- 12) Prepositional periphrastic (Prep + VN) --- NO [MDM, p.c.]

- 13) "DO" periphrastic ("DO" + VN) --- YES, in a sense
 --- Compound verbs very common, especially with nativized Arabic lexemes; syntagm is: X + Simple Verb, where Simple Verb is drawn from a set of some 14 basic verbs, and

X = {N, Adj, Adv, PrepPhr} (85);

X can be a "Verbal Noun" (86) in the derivational/nominal sense described above (see [10]), notably the form in -eŝ which can have verbal-abstract semantics; this use with -eŝ is productive, often tending to supplant the older simplex verb [MDM, p.c.]; e.g. (86)

kušidan "try" ==> kuš-eš "trying" ==> kuš-eš kardan "try"
trying do

--- For some compound verbs, do have X = Present Stem of a verb (or some other part of the verb) (89); this fits the syntagm, but lexically restricted

- 14) Adverbial clause = "and" + finite clause --- not exactly
--- Per Lambton (139), conjunction va, o "and" may be used "to introduce a qualifying phrase", ex.

at table sitting he-was and pen-a in hand-his it-was
 "He was sitting at the table with a pen in his hand" ;

but oddly translated, really more like two separate events [MDM, p.c.]

- 15) VN/Inf instead of finite main-clause form --- weakly

--- If verb in successive coordinate clauses takes the identical Aux element, may omit Aux from all except first or last clause (162); see this as Gapping [MDM, p.c.]

--- Per grammar, a compound verb in first of two coordinate clauses may omit its inflected verb component (162), regardless of whether second verb is compound; but construction extremely strained, because no real Gapping [MDM, p.c.]

--- NB: tenses built with Aux involve either past participle or "short infinitive" (17-18), neither of which is the "Inf" proper; nor are compound verbs formed from "Inf" (89)

- 16) Word-initial change --- no mention, assume NO

- 17) Extended use of kin terms ("Kin of Noun") --- NO

--- No examples in Aryanpur-Kashani & Aryanpur-Kashani 1978 (large dictionary); MDM tentatively confirms

Pipil (Mesoamerica; Aztecan [Uto-Aztecan])
Campbell 1985

- 0) Basic categorial information

--- Noun: no cases ("absolute" suffix not a case marker, 39); no gender mentioned; nouns code plural (51-53), take possessive prefixes (42)

- Verb: Subj and Obj person/number coded affixally (54-55)
 - Article: preposed definite and indefinite articles (56-57)
- 1) Conjugated adpositions --- YES (but only for relational nouns)
 - Relational nouns (59-60), analyze here as "hedged prepositions"; pattern like possessed NPs ("my-with" = "with me"), take possessive pronominal prefixes (42ff., 56); full-NP object takes a proleptic pronominal prefix:
 - 3poss-Adp DeptN "its-under [the tree]" = "under the tree"
 - There are some "independent prepositions" (59), apparently used only with full NPs and not taking proleptic pronouns (?); no postpositions (61)
 - 2) Word order
 - VOS (102)
 - Prep-Obj (59-61); same order with relational nouns (59)
 - Adj-N (120); rare N-Adj, especially when Adj is Spanish loan
 - N-Gen (117)
 - N-RC1 (128-30); order not stated as such but inferrable from exx. and discussion
 - 3) Relative clause linker: INVARIANT (choice of several) (128)
 - 4) Relativization strategy/ies
 - Syntagm: Head + linker + normal finite clause (128); apparently can only relativize Subj and Obj, not Possessive or Obj-of-Adpos (129-30)
 - Coreferential N is itself deleted; but verb always co-indexes Subj and Obj even in main clauses (54-56), therefore also in RC1s
 - 5) Special relative form of verb --- NO
 - 6) Polypersonal verb --- 2 affixal arguments, Subj and Obj (54-55)
 - Verb can take only a single Obj marker (109)
 - 7) Infixing/suffixing alternation --- NO
 - Order of affixes is always: Subj-Obj-V; language has Aux elements, but don't affect Subj/Obj placement
 - 8) Definite article in genitive embeddings
 - Genitive syntagm is: Poss-N Gen ("his-house the man"), or N pal Gen (117, 118)
 - Articles (56-57) can occur on both Nouns, or just on Gen (not stated explicitly, but clear from exx.) (117-18):
 - (Art)-Poss-N Art-Gen
 - 9) Nonconcord of V with full-NP Subj --- NO
 - Verb regularly cross-references full-NP Subj and Obj (74); nothing more is said
 - 10) Verbal abstract: VN or Inf? --- NEITHER
 - Language uses fully finite verbs in Equi contexts (140-42) and for purpose clauses (134-35)

- 11) Predicative particle? --- NO
 --- Zero copula; order: Subj Pred ("I [am] the king"); no particle (108, 111)
- 12) Prepositional periphrastic (Prep + VN) --- not applicable
- 13) "DO" periphrastic ("DO" + VN) --- in a sense
 --- Only with Spanish infinitives (used as nouns) (143)
- 14) Adverbial clause = "and" + finite clause --- no mention, assume NO
 --- Adv clauses take their own conjunctions (130ff.); there is a clause-level "and" word (121-23), but not mentioned in adverbial use
- 15) VN/Inf instead of finite main-clause form --- not applicable
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- NO
 --- Dictionary gives only "child moon" (= "new moon"), more like a diminutive semantically; apparently compound, not possessive

Quechua (Imbabura) (Ecuador; Andean)
 Cole 1982

- 0) Basic categorial information
 --- Noun: case (expressed solely through postpositions, 76); plural suffix (128), usually obligatory; no gender (129); no article (77, 162); no possessive suffixes (115; found elsewhere in Quechua, but lost in Ecuador)
 --- Verb: codes Subj person/number suffixally (obligatory), with no number distinction in 3rd person; also 1-sg Object (NB: only 1-sg; optional) (159-60)
- 1) Conjugated adpositions --- NO (162)
 --- Grammar equates "Postpositions" and "case endings" (76); analyze here as "hedged postpositions"
 --- No affixal pronouns, only free (a single undifferentiated series) (129-30); postpositions attach identically to nouns and pronouns (129, exx. 104, 115)
- 2) Word order
 --- SOV, with flexibility in main clauses but rigidly OV in subordinate clauses (71)
 --- Postpositions (75), no Prepositions (162)
 --- Adj-N (77) (Adj distinguished from N syntactically but not morphologically, 99, 186); modifiers precede the N (76)
 --- Gen-N (77, 115)
 --- RCl-N, or internally headed (49, 77)

- 3) Relative clause linker --- ZERO
- 4) Relativization strategy/ies --- GAPPING (with external head)
 --- Internally headed (RC1-N), or externally headed, or extraposed (N ... RC1) (50-51)
 --- RC1s are non-finite (no person-number marking, owing to disappearance in Ecuadorian Quechua of possessive suffixes, 33-34); verb-form is followed by nominalizer suffixes (47) (also used to form Noun clauses); nominalizers indicate tense, minimally also role of relativized noun (47, 50, 175)
- Externally-headed type (RC1-N): Only HeadN is case-marked (47ff.)
 Internally-headed type: Only nominalized V is case-marked
 Extraposed type: Both are case-marked
- Can relativize on Subj, Obj, Obj-of-Postp, Adv (53-54), but not Possessor (58); in externally headed RC1, obligatorily gap coref Noun together with (NB) its Postposition (53); recall that Postp is hedged
 --- DObj within RC1 often incorporated into the V, i.e. loses Acc case-mark (48)
- 5) Special relative form of verb --- not exactly
 --- Nominalizer suffixes (47) also used to make Noun clauses
- 6) Polypersonal verb --- codes Subj and 1-sg Obj suffixally (159)
 --- 1-sg Obj may be either suffix or free pronoun or both (103-4)
 --- Non-Ecuadorian Quechua codes both 1- and 2-person Objects, via complex conflated Subj/Obj marking (159-60)
- 7) Infixing/suffixing alternation --- NO
 --- Various compound verb forms involve an Aux ("be", see [12] below), but Aux is a fully inflecting verb; no mention of whether 1-sg Obj affix moves (assume it stays attached to transitive main verb, not to intransitive Aux "be")
- 8) Definite article in genitive embeddings --- not applicable
 --- Possessor NP takes "possessive" case marker -paj: PossNP-paj HeadN; construction identical for Noun and Pron possessor, though -paj often omissible with 1-sg "my" (115)
 --- No article
- 9) Nonconcord of V with full-NP Subj --- NO
 --- Subject-V agreement obligatory in matrix clauses (159), but 3-person Subj-markers do not distinguish sg/pl anyway (160; marker is zero in some tenses, non-zero in others, 143-45)
- 10) Verbal abstract: VN or Inf? --- INFINITIVE
 --- Exx. passim show normal marking (accusative -ta) of Obj of non-finite verb forms; Object N may be incorporated into the V (36-37, 48), as signalled by loss of Acc case, but no genitive marker -paj involved
- 11) Predicative particle? --- sort of
 --- For both NounPred and AdjPred, the usual construction is:
 Subj(Nom) Pred(Nom) Copula (67, 103, 105)

--- Copula (ka-) obligatory except in 3-person present, when usually omitted (67)

--- Sentences often take one of various "validator" morphemes (= evidentials, 164ff.), which appears encliticized to the focus/rheme of the sentence (165); naturally, rheme in copular constructions is the Pred; and validator is obligatory if copula omitted (67):

Subj Pred-Validator "Subject is Predicate"

--- So might take the validator as being a "predicative particle"

12) Postpositional periphrastic (VN + Postp) --- no mention, assume NO

--- Various compound verb forms made with "be" ("attitudinal" aspect (148-51), past conditional (155-56), passive (133)), but no postposition involved; recall that Postp is hedged

13) "DO" periphrastic (VN + "DO") --- no mention, assume NO

14) Adverbial clause = "and" + finite clause --- NO

--- There are various clause-level "and" elements (78); but no hint of such usage

--- "Adverb clauses ... are non-finite" (60); "The use of adverbial subordination in place of coordination is typical of the Quechua languages" (79), i.e., just the inverse of our construction

15) VN/Inf instead of finite main-clause form --- NO (158)

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- not really (222-23)

--- Two exx. seem clearly augmentative: "mother finger" = thumb, "mother tooth" = molar; also "baby mother" = placenta; semantics not right, and also not genitives but compounds (198)

Shoshone (Panamint) (SE California; Central Numic [Uto-Aztecan])
Dayley 1989a

0) Basic categorial information

--- Noun: 3 cases (Nominative, Objective, Possessive) (176ff., 185-86); no articles; no gender; number marking optional (216)

--- Verb: inflects for number (obligatorily, 77), but not person (53)

--- Article: none described; but demonstratives (which may be used either as pronouns or as determiners) often function much like English article (138); exx. seem uncommon in grammar; will not reanalyze as "hedged article"

1) Conjugated adpositions --- NO

--- Postpositions (190ff.) take Obj-case pronouns (207), which are presented as independent pronouns (129-30); however, I find (twice) the word umma [= u + ma], presented as a fused form and glossed "it-with" (exx. 369, 371)

- 2) Word order (14-15)
 - SOV (20-21); OV holds overwhelmingly, SV is only a weak trend
 - Postpositions (190ff.)
 - Adj-N (272-73); N-Adj possible, but here Adj analyzed as RC1 (274)
 - Gen-N (186)
 - N-RC1; RC1 can be separated from Head-N and dislocated to end of matrix clause; also RC1-N (365-66)
- 3) Relative clause linker
 - "Obviative demonstratives" (135ff.) function as "Relative Pronouns" (157-58), take case-marking per matrix clause (357-58); RelPron optional (362), may precede or follow RC1 (not necessarily coming between Head and Clause), even occur twice (before and after, 363-65)
 - One ex. (explicitly mentioned as such) shows RelPron marking case within relative clause, not matrix clause (364); author admits RC1s not wholly understood (365)
- 4) Relativization strategy/ies --- GAPPING; Copy/Correlative (Postp)
 - Verbs in RC1 (358) occur in 3 nonfinite forms (Inf, PresentPtcpl, PastPtcpl, see 122ff.), depending on tense and on Subj vs. Non-Subj RC1; usually N-RC1, also RC1-N
 - For all non-Subj RC1 (including Postp): Subj of RC1 assumes Possessive case (per exx. 358ff.; no explicit statement), and assumes special possessive-reflexive marking if coreferential to Subj of matrix clause (361-62)
 - Postpositional RC1 ("Oblique RC1", 368): construction quite different, apparently does not use RelPron but normal personal pronoun as Obj-of-Postp; PostpPhr [Pron+Postp] fronted in RC1 (NB: not movement to verb), and RC1 itself usually fronted before MainCl (which contains HeadN, resuming pronominal Obj-of-Postp); this looks like correlative type, with pronominal copy in RC1 (not MainCl); but verb still nonfinite and Subj still in Possessive case; analysis unclear
 - No mention of genitival RC1, assume not possible; no Possessive-case form of RelPron mentioned (paradigm, 357)
- 5) Special relative form of verb --- not exactly
 - Infinitives and participles have other subordinating uses (122-25): PresPtcpl also for AdvCl (356), Inf and PastPtcpl for NounCl (374)
- 6) Polypersonal verb
 - Verb does not code person at all (53)
 - Many verbs (but not all, 76) code number of their S or O argument (i.e. ergatively) via stem suppletion, Ablaut, or plural suffix (72-75)
 - Also, Imperative obligatorily codes number of Subj via an enclitic, which occasionally appears on other forms (78-79); thus sometimes can code number of both Subj and Obj (Imperative exx., 79)
- 7) Infixing/suffixing alternation --- NO
 - If number coding (of Obj) involves an affix at all, it's strictly suffixal
- 8) Definite article in genitive embeddings --- not applicable
 - No article; syntagm is: DeptN HeadN
 - Examples of Gen-N embedding (186-87) show one ex. with a

demonstrative, occurring on DeptN:

that cow's head meat "the meat of the cow's head" (187)

9) Nonconcord of V with full-NP Subj --- NO

--- Concord only in number, but "not dependent on overt marking of number" on Subj or Obj NP (77); V must mark number, though nonhuman NP need not (77)

10) Verbal abstract: VN or Inf? --- assume INFINITIVE

--- Category called "Infinitive" (ending -nna), functions as infinitive or gerund (122-23), used in EQUI (381-82) and purpose clauses (355-56), also in non-Subj RCLs (358)

--- Various other "subordinate" verb-forms partially overlap functionally with Infinitive (344-47); no explicit statement in grammar re case of Obj with any of these (nor with Inf), and hardly any relevant exx.; object seems never to appear in Possessive case (per ex. p. 381, also see 356)

--- In some constructions, subjectless Inf must be in passive (374-75), hence object does not surface as such

11) Predicative particle? --- NO

--- Predicate nominals in Nominative case, no particle (27ff., 177); stative verb naa as "copula par excellence" in equational sentences, but optional in Present (27ff., 277ff.); various word-orders passim, but basic syntagm apparently:

Subj Pred (Cop)

12) Postpositional periphrastic (VN + Postp) --- NO

--- The only Postp that can be used with verbs at all is "like" (197), adding nuance "might"; no copula involved (cf. [13] below)

13) "DO" periphrastic (VN + "DO") --- barely?

--- Aux verbs exist (80ff.), but no DO-Aux

--- But note the usage in [12] above (only one ex., 197):

those drinking like do-durative "They seem to be drinking"
(lit.) "They're doing drinking-like"

14) Adverbial clause = "and" + finite clause --- no mention, assume NO

--- Coordination (337ff.) vs. Adverb-clause subordination (347ff.), no overlap; sentence-level "and" conjunction exists (339)

15) VN/Inf instead of finite main-clause form --- no mention, assume NO

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- NO

--- No exx. in Dayley 1989b

Slave (Northwest Canada; Athapaskan)

Rice 1989

[My thanks to Keren Rice for answers to several queries]

0) Basic categorial information

--- Noun: no cases; 3 genders (general, wooden, areal), not marked on N itself (1023ff.); optional plural -ke for human nouns (247); possessive prefixes (157, 207ff.)

--- Verb: Subj and Obj person/number coded prefixally (425ff., 1005ff.); verb also can incorporate Postpos and its object (425); gender can be coded for some verb themes (1023)

--- Article: not really; demonstrative pronoun ?eyi "this, that" (261) can occur prenominally as determiner, is sometimes translated "the" *pas-sim* (ex. 1001); often used with Head-N of RC1 (1310); will analyze as "hedged article"

1) Conjugated adpositions (269ff.) --- YES

--- A single prefix series serves for pronominal Obj-of-Postp, Possessor-of-N, and DObj (627)

--- Normally, a full-NP Obj-of-Postp excludes a cooccurring pronominal object (269)

2) Word order (997ff.)

--- SOV (997)

--- Postpositions (1004)

--- Gen-N (1001, 1004)

--- N-RC1 or internally headed (1309)

--- No real "adjectives" (1004); there is a category "Adjective", but occurring only as complement of V, notably "be" (390); many adjectival notions expressed by stative verbs (576ff., 908)

3) Relative clause linker --- invariant or zero

--- Various complementizers occur after the RC1 (either externally or internally headed) (1310, 1315ff.); not positioned right to be "linkers"

--- Invariant "relative pronoun" t'aa ("unspecified in reference") may optionally follow HeadN; grammar gives exx., discusses difficulty of determining whether "RelPron" falls within matrix or embedded sentence (1324-25)

4) Relativization strategy/ies --- 1) Internal head; 2) Copy, Gap/Copy

--- (1) Internally headed (1310ff.): RC1 seems identical to an independent sentence; coreferential (internal-head) noun apparently not singled out in any way (possibility of ambiguity, 1311-12)

(2) Externally headed (1313ff.): order is N-RC1; RC1 includes a pronominal copy of the head (marked on V or Postp in all exx.; also apparently can be pronominal Possessor [KR, p.c.])

(3) Subject RC1 (1312-13): choice of analysis (1) vs. (2) unclear

--- All RC1s followed by complementizer (on RC1-final V) (1331); HeadN often preceded by determiner ?eyi "that" (1310); linker may separate HeadN and following clause

--- Can relativize on embedded Subj, DObj, Obj-of-Postp (1315); also on Possessor and Object-of-Comparative (KR, p.c.; exx. in grammar don't make the intended point, 1315); for externally-headed RC1, Postp RC1

counts as Copying (see [1] above), Poss RCl counts as either Copying or Gap/Copy, depending on alienability (see [8] below)

5) Special relative form of verb --- NO

6) Polypersonal verb (425ff., 1005ff.) --- 3 actants coded

--- Slots for Subj, DObj; verb also can incorporate Postpos and its Obj (433, 741ff., 775-76; Postp can be zero, 746ff.), thus adding an oblique actant (only one [KR, p.c.])

7) Infixing/suffixing alternation --- NO

--- Position of elements fixed in slot sequence (425)

8) Definite article in genitive embeddings --- hedged

--- Order is Gen-N (1001); have both alienable possession (suffix *-e* on HeadN) and inalienable (suffix floating high tone on HeadN) (212-13); inalienable requires pronominal-possessor prefix even with full-NP possessor, alienable usually forbids it (228-29)

--- In Gen-N constructions, Dept can take a determiner, Head cannot (1001); example (with "article" *?eyi*) is glossed

[the boy] dog = "the boy's dog" (order: [Det Gen] Head)

9) Nonconcord of V with full-NP Subj

--- When independent subject pronouns occur, verb shows normal person/number concord (per exx. 253-54, 1201)

--- 3rd-person Subj marker on verb (slot 12) is always zero, so no change with full-NP subject (1017)

--- The separate plural-subject morpheme on V (slot 7) offers possibilities for (non-)concord: if Subj is human plural, either V or N or both is marked with plural *-ke* (248, 1017); hence optional concord

--- As for Obj, pronominal/affixal Obj (marked on verb) and full-NP Obj are normally mutually exclusive (1017), i.e. concord is not an issue; but if full-NP Obj is human plural, V normally does code 3pl object (in normal Obj slot) (1017)

10) Verbal abstract: VN or Inf? --- NEITHER

--- All complementation seems to involve finite clauses (1221ff.; exx. involve prototypical EQUI-type verbs)

--- "Deverbal nouns" (170ff.) apparently involve concrete notions (Agent, Instr, etc.); derived nominals, not inflectional verbal abstracts

11) Predicative particle? --- NO [KR, p.c.]

--- Adjectival predicates are either (1) "Adjectives" accompanying a copula-like verb (389), or (2) stative verbs (908); in neither case is there a particle

--- Noun predicates not discussed in grammar; per KR [p.c.], construction has no particle, follows pattern: Subj NPred Copula (misc. exx. 393, 396, 1301; see 1299-1301 for copulas)

--- Some nouns are directly usable as predicates ("denominal verbs", 933)

12) Postpositional periphrastic (VN + Postp) --- not applicable

- 13) "DO" periphrastic (VN + "DO") --- not applicable
- 14) Adverbial clause = "and" + finite clause --- apparently NO
--- Sentential "and" discussed (1049-50), but no adverb-clause use mentioned
- 15) VN/Inf instead of finite main-clause form --- not applicable
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- NO [KR, p.c.]

Squamish (Pacific NW coast; Salishan)
Kuipers 1967

0) Basic categorial information

--- Noun: two cases, Absolutive (zero marker) and Relative (t- prefix on Article) (136), where Absolutive is used for Subj and DObj, Relative otherwise; no gender mentioned; plural expressed (optionally, 100) by reduplication (98ff.); possessive affixes (87)

--- Verb: Subj and Obj person/number marking essentially suffixal (85ff.), but certain proclitics may attract Subj markers (see [7] below); verb conjugates in four paradigms: Nominal, Finite, Hypothetical, Factual (87-88); Subj coding in Factual paradigm involves possessive prefixes as well as Subj affixes (91)

--- Article: both definite and indefinite (137ff.), proclitic; semantically not quite like English

1) Conjugated adpositions --- not applicable

--- No category "adposition"; expressed by "relator verbs" (153), e.g. "be-at"

2) Word order

--- VSO (169-70)

--- N-Gen (176)

--- N-RC1 or RC1-N (175-76); grammar presents N-RC1 and RC1-N as instances of "apposition" and "attribution", respectively, but apparently these terms are employed (definitionally) merely as labels for the different word orders (?)

--- No categories of adposition, adjective; Adj expressed by stative verbs

3) Relative clause linker

--- Subj- and Obj-relatives: zero linker; Oblique RC1s can be preceded by t1- (161)

4) Relativization strategy/ies

--- Verbs have two special "Nominal paradigms" (88), expressing (headless) Subj-Rel and Obj-Rel; Subj-Rel verb form ("one who helps (me)")

has no Subj-affix, Obj-Rel verb form ("one whom I help") has no Obj-affix; but 3-sg Obj affix is always zero anyway (85), while 3-sg Subj affix is /-as/ in some paradigms but zero in others (86); so impossible to tell if an argument has been "Gapped" in the verb morphology in these RCLs or not; since construction is inherently headless, might speak of full-NP as "gapped", but would seem odd

--- Headed Subj- or Obj-RCLs: Head-N is either preceded or followed by a Nominal-form verb (175-76)

--- Note the construction (177)

a man [big his nose] = "a man with a big nose" ;
here the resumptive pronoun is apparently present even in non-RCL (see [8] below), thus Gap/Copy type; this construction is the only indication of a genitival RCL in the grammar, and will be counted here as such

--- Oblique RCL can be expressed by RCL where the (notional) prepositional idea is expressed by a "Relator verb" and the RCL is juxtaposed to the Head: "the-box [BE.IN the-clothes]" = "the box in which the clothes were" (177); adverbial argument gapped

--- Also have a special clause type introduced by morpheme tl (161, 196ff.), meaning "why, where, when, how, ... about whom, with which, etc." (196); may be used for Oblique RCLs (198), e.g. "the-barrel tl-BE.IN the-molasses" = the barrel in which the molasses were"; adverbial argument gapped; unclear what form the V assumes, but apparently not the Nominal form (183)

5) Special relative form of verb --- YES (non-Oblique)

--- Nominal verb-form series used expressly for Subj- and Obj-RCL

6) Polypersonal verb --- 2 actants coded on verb

--- V-Obj-Subj (85); ordering differs if V preceded by clitic č- (see [7] below); "Factual" paradigm (90-92) incorporates Possessive pronouns into paradigm in complex fashion

7) Infixing/suffixing alternation --- YES, but for Subj (!)

--- 1st/2nd-person forms in "Finite" paradigm are preceded by clitic č- (89, 156-57), followed by a Subject-suffix; order is thus:

č-Subj-V-Obj

--- 3rd-person "Finite" forms usually precede by clitic na (89, 157); but na does NOT attract Subj marker (or Subj marker is zero)

--- Proclitics q (189) and tl (196) can also attract Subj markers, with the latter now appearing not as affixes but as "personal clitics"

--- Alternations affect only Subj suffixes; Obj always suffixed to V

8) Definite article in genitive embeddings

--- Article freely appears on both N and Gen; either (176):

(a) Art-N Art-Gen: the-house the-Tom = "Tom's house"

(b) Art-N-his Art-Gen: the-house-his the-chief = "the chief's house"

--- Pattern (a) for proper name, (b) for common noun; article on Gen is in Relative case in (a), Absolutive case in (b)

9) Nonconcord of V with full-NP Subj --- YES (optional)

--- For 3-pers verb with plural Subj, plural "is often not expressed [on verb] when the subject-noun immediately follows" (86)

- 10) Verbal abstract: VN or Inf? --- analyze as NEITHER
 --- Verb-nominalizer s-; occurs both (a) in fixed lexicalizations, and (b) freely combining with any verbal stem to refer to a fact (66-67), in the "Factual" verbal paradigm (90-92); exx. (183ff.) show this used in prototypical infinitival contexts
 --- Subject does continue to be expressed, via complex combinations of Poss and Subj markers (Factual paradigm)
 --- Obj continues to behave like Obj (though grammar does not state this): continues to take Obj-affixes on verb (91, in "Factual" paradigm), and see ex. on p. 185 ("I want to learn the Squamish language") where Obj NP ("the Squamish language") takes article in Absolutive case
 --- Would count unambiguously as nominalized finite clause, except that expression of Subj partly involves Poss suffixes; analyze here as finite (perhaps with some voice change)
- 11) Predicative particle? --- NO
 --- Predicate nominals occur initially in clause (169), no particle
- 12) Prepositional periphrastic (Prep + VN) --- not applicable
 --- But NB: the clitic na appearing with some 3rd-person verb forms is etymologically identical to the Relator-verb na? "be-at" (functionally similar to Preposition, see [1] above) (157)
- 13) "DO" periphrastic ("DO" + VN) --- not applicable
 --- But NB: the clitic č- appearing with 1st/2nd-person verb forms is a reduced form of ča(?) "do, act, make" (156)
- 14) Adverbial clause = "and" + finite clause --- YES
 --- Conjunction "and" (mentioned in chapter on clausal Coordination) also translates as "while, so that, for, until, but" (214)
- 15) VN/Inf instead of finite main-clause form --- not applicable
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- NO
 --- Apparently not; checked terms for "child, father, grandchild" in glossary at back of grammar book (254, 255, 398), without result

Sumerian (ancient Mesopotamia; isolate)
 Thomsen 1984; also Falkenstein 1959, Gragg 1972
 [My thanks to Dan Foxvog for advice and clarification]

- 0) Basic categorial information: agglutinating
 --- Noun (summary 48): suffixal marking of pron-possessor (71), plural (59-63), case (88); anim/inanim gender (49); no articles; various kinds of plurality marked by suffixes or reduplication (59-63); case system is Erg/Abs (49), with the cases termed "postpositions"; order is:
 N - possSfx - pl - case (54) ;

case ending is added (once only) to NP as a whole, following any post-nominal modifier, viz. Adj (64), Gen (91), RC1 (242)

--- Verb (summary 50-51): verb marks pers/numb of Subj and Obj using both a prefix slot and a suffix slot (148-49, 152); which slot codes which role varies, depending on perfective (hamtu) vs. imperfective (maru) Aspect (117, 141-44), and sometimes a pers/numb marker may straddle both slots; verb may also incorporate one or more Postps (sometimes in an altered form), often with their pronominal Obj-of-Postp ("dimensional infixes", 214ff.); order is:

[(pron) + postp] - pfx - root - sfx (139) ;

also 7 verbs have suppletive sg/pl roots, usually reflecting plurality of the Abs participant (131-32)

--- Caveat: much of our knowledge of the grammar still tentative, especially since early Sumerian tended not to write grammatical morphemes (20ff.)

1) Conjugated adpositions --- none or barely

--- "Postposition" equated to "case marker" (88); ten cases, no other Postps; analyze here as "hedged postpositions"; Postp may be incorporated into verb

--- Combination [Pron+Postp] apparently involves no phonological fusion, though note a few puzzling forms from Gudea (68) [also DF, p.c.]

2) Word order

--- SOV (51)

--- Postpositions (88)

--- N-Adj (54, 63-64)

--- N-Gen, also Gen-N (90-91); see [8]

--- N-RC1 (242)

3) Relative clause linker --- ZERO, or showing gender only

--- Optional linker: use the word "man" as linker for animates, "thing" for inanimates (242); but odd example (250): "his throne which he has erected", with "which" rendered as "man"

4) Relativization strategy/ies

--- Basic syntagm (242ff.): HeadN (Linker) [RC1...Verb]-a-case, where "case" = case of HeadN, and -a- is multi-purpose "subordinating suffix" (241)

--- Little information; Gragg (1972:154) says that Subj and Obj RC1s "can leave no pronominal trace", whereas genitive and oblique RC1s may leave resumptive pronominal copy if HeadN is animate (NB: he does not say "must"); by contrast, oblique RC1s with inanim HeadN involve gapping (155); but Gragg's examples use invented data, and the constructions feel odd [DF, p.c.]; might be likelier to resume Postp Obj inside the verb as incorporated PostpPhr, but even this is speculative [DF, p.c.]

--- Also can make RC1s via nonfinite form in -a (see [10] for morphology), acting participially; identical form may serve as intrans ptcpl, or passive trans ptcpl (with Erg argument either retained or postposed as genitive after the nonfinite verb), or active trans ptcpl (1984:261-63):

sun [rise-a] = the rising sun
 young.man [An-erg call-a] = the young man called by An
 " [call-a] An-gen = " " " " " "
 man [slave buy-a] = the man who bought the slave ;

construction may also be used for HeadN functioning as Oblique in RC1 (263)

5) Special relative form of verb --- YES, sometimes

--- Finite -a form also used as complement of verb (usually verb of saying) (241)

--- Non-finite -a form apparently restricted to RC1 function; a distinct but related nonfinite form is used as complement of verb (261)

--- One verb (duq₄) has a special suppletive nonfinite form di (301 [also DF, p.c.])

6) Polypersonal verb

--- Verb codes Subj and Obj affixally (see [0]); also may have one or several (rarely more than two) incorporated Postp phrases with pronom Obj

--- As a rule, independent pronouns have no Obj form, hence Obj must be expressed in verb (69)

7) Infixing/suffixing alternation --- weakly

--- Obj appears as suffix or as prefix, depending on Aspect (see [0]); this "prefix" is seldom truly initial in verb slot-sequence, i.e. more like infix; however, alternation does not depend on presence/absence of a preverb

8) Definite article in genitive embeddings --- not applicable

--- No definite article; syntagm (90-91):

[Head Dept-gen]-case (where "case" = case of Head)

or: [Dept-gen Head-poss]-case ("anticipatory genitive")

9) Nonconcord of V with full-NP Subj --- unclear

--- Only animate nouns can have true suffixal plurals (59), hence concord is basically an issue only for animates; per Falkenstein, an animate plural noun ("Personenklasse") may sometimes take 3-sg-inanim verb, used collectively (1959:58)

--- Nouns (especially inanimates) may reduplicate (61), as may verbs (123), with various nuances expressed including (sometimes) plurality [DF, p.c.]; unclear what kind of Subj-Verb "concord" this quasi-plurality displays

10) Verbal abstract: VN or Inf? --- INF

--- Verbal abstract formed by deleting all prefixes and all personal affixes, usually adding subordinating affix -a (1984:254); three basic types, of which (i) shouldn't count as verbal abstract:

- (i) Hamtu-a --- only "participial", adnominal (RC1) (261)
- (ii) Marû-ed-a --- complement of verb ("Subjunctive") (265)
- (iii) Marû-ed-e --- intentional/future adverbial (266)

--- Falkenstein explicitly states that nonfinite forms have normal verbal rection (1959:57); no statement in Thomsen, but all arguments of "Inf" remain preverbal (exx. passim), with noun Obj in Abs case

--- Alternate syntagm whereby ergative Subj follows nonfinite verb-form, in genitive case (1984:92, 262, see [4]); but Obj still in Abs

11) Predicative particle? --- NO

--- Enclitic copula (275-78), syntagm: Subj Pred-Cop;
Cop inflects for pers/numb; may also occur after finite verb in emphasizing function (277)

12) Postpositional periphrastic (VN + Postp) --- no mention, assume NO

--- No mention under uses of nonfinite forms (255ff.) or of Postps (92ff.); recall that Postp is hedged

13) "DO" periphrastic (VN + "DO") --- sometimes

--- Lexicalized compound verbs exist, with Obj+Verb comprising a semantic unit (269); Obj apparently is a normal noun

--- However, entire compound verb itself may stand as Obj of Aux verb "do"; no apparent semantic difference between regular compound verb and such "double compounds"; construction appears to exist only for compound verbs (271)

14) Adverbial clause = "and" + finite clause --- apparently NO

--- Only "and" conjunction is u, borrowed from Akkadian (83); usual nuance "and then, moreover, but", not merely coordination

--- Adv clauses usually formally RCLs with temporal/locative HeadN (246), or headless nonfinite Hamtu-form with possessive suffix (264)

15) VN/Inf instead of finite main-clause form --- apparently NO

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- not really [DF, p.c.]

--- DUMU.URU "child city" = citizen; in the example AMA.BAPPIR (mother.beerbread) = agarin "a baked good", the semantics aren't completely clear, and anyway only the written form (AMA.BAPPIR, composite logogram) is at issue

Tagalog (Philippines; Northwest Austronesian)
Schachter & Otnes 1972

0) Basic categorial information

--- Noun: no cases; no gender; plural coded (optionally) with proclitic mga (111ff.); no articles

--- Verb: no affixal person inflection; verb optionally coded for plural (335)

1) Conjugated adpositions --- not applicable (?)

--- Conceivably could call the pronominal case markers "prepositions"; otherwise category doesn't seem to exist

--- ng-phrase (Actor), ang-phrase (Topic), sa-phrase (Directional) each have their own suppletive pronominal paradigms (88); ng- and ang-series pronouns are free-standing, sa-series pronouns are preceded by sa (but no fusion or other evidence for "conjugated preposition"); para "like" links to its object with linker -ng (253)

2) Word order

--- V-first (66)
 --- Prepositions (if category exists)
 --- N-Adj or Adj-N, freely (122)
 --- N-Gen or Gen-N (134ff.); several strategies, one (more or less) strictly N-Gen, others flexible (see [8] below)
 --- N-RC1 or RC1-N (the former preferred if RC1 is long) (123)

3) Relative clause linker --- INVARIANT (na/-ng) (123)

--- Same linker, in either order: N Linker RC1 ~ RC1 Linker N

4) Relativization strategy/ies --- GAPPING (123)

--- RC1 identical to main clause except for gap of coreferential N; this gapped N must be Topic in RC1, thus apparently excluding genitival RC1

--- Also can build headless RC1s by nominalizations (150ff.); in some such cases, Object complement may change "case" from ng (Object) to sa (Directional) (382)

--- Note construction "child [new (is) TOPIC-pencil]" (135-36), roughly translatable as "child with the new pencil, child whose pencil is new"; this will not be construed as a genitival RC1

5) Special relative form of verb --- NO

6) Polypersonal verb --- one actant coded (number only)

--- Number of Topic optionally coded in verb (335-36); no person marking at all

--- Verb arguments are either noun or pronoun (exx. passim); argument may be optional, depending on focus-type of V (73)

--- Inasmuch as some personal pronouns are constrained to 2nd position (i.e., postverbal), they are "enclitic" to verb (183ff.); but "enclitic" is defined (183) in purely positional terms, with no indication of any phonological fusion of verb and pronoun; most pronouns disyllabic (88); apparently pronouns don't cooccur with full-NP Subj; thus pronouns should not count as verbal pers/numb markers

7) Infixing/suffixing alternation --- not applicable

8) Definite article in genitive embeddings --- not applicable

--- No article; various genitive constructions (134ff.), one with pre- and post-Linker elements reversible (a), one with order more or less fixed (b):

- (a) HeadN Linker [may/sa DeptN] (135), where may or sa is possessive predicating element in "have" sentences (273ff.)
- (b) HeadN ng DeptN (136)

- 9) Nonconcord of V with full-NP Subj --- sort of
 --- Number is coded (optionally) on Nouns (111) and Verbs (335); plural can usually be marked on verb or Topic or both (111), thus agreement normally optional
- 10) Verbal abstract: VN or Inf? --- INFINITIVE
 --- Several ways to create gerund-like verbal abstract (154ff., 159ff., 164ff.); for all these, grammar states that the verbal abstract may be followed by any of the normal objects or complements (in their normal form) (155, 163, 165)
 --- Also find finite clauses as object complement, etc. (177ff.)
- 11) Predicative particle? --- NO
 --- Equational sentences lack copula; Pred comes initially (61)
- 12) Prepositional periphrastic (Prep + VN) --- not applicable
- 13) "DO" periphrastic ("DO" + VN) --- no mention, assume NO
- 14) Adverbial clause = "and" + finite clause --- no mention, assume NO
 --- at "and" can introduce complement clauses in the frame
 "It's good / I'm glad / Too bad) that (= "and") ..." (545-46)
 --- No hint of "and" to introduce real adverbial clauses (463ff.)
- 15) VN/Inf instead of finite main-clause form --- no mention, assume NO
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- a little
 --- "Child of the sun" = albino (perhaps calque on Spanish "hijo del sol" = albino; cf. Campbell's Pipil grammar, 1985:706); "child of sweat" = laborer; "mother of stairs" = side-boards of a staircase (to which the steps are nailed); "child of wealth" = someone born wealthy; "father of city" = mayor; a few others of this sort

Tamil (southern India; Dravidian)
 Lehmann 1989 (Modern Literary); and Asher 1985 (colloquial)
 [My thanks to Susan Herring for answers to several queries]

- 0) Basic categorial information: Agglutinating (8)
 --- Noun: number and case coded suffixally (11, 13); Nom case may be used for DObj, esp. if noun is non-rational and/or indefinite/generic (26ff.); some cases expressed by suffixes, some by "bound postpositions" (23-24)
 --- Verb: finite verb marks pers/num/gender of Subj only (suffixally) (48)
 --- Article: "one" used as indef article (112); no definite article (Asher 61)

- 1) Conjugated adpositions --- NO
 --- "All pronouns are free forms" (Asher 141)
- 2) Word order
 --- V-final, nouns permute freely (176); SOV clearly preferred (Asher 55)
 --- Postpositions (117)
 --- Adj-N (134)
 --- Gen-N (Asher 113)
 --- RC1-N (250-51, 285), or correlative (RC1-MainCl, 349ff.)
- 3) Relative clause linker --- ZERO
- 4) Relativization strategy/ies --- GAPPING (participial); correlative
 --- External-head strategy (286ff.): clause-final V becomes "Adj.Ptcpl" (tense is coded, but in positive verbs only; Subj not coded); coref N gapped (with its case-mark)
 --- This strategy allows relativization on Subj, DObj, IObj, but only some obliquely case-marked NPs; can relativize on certain Postps, stranding the Postp (but cf. Asher 30: can relativize freely on Postps), but not on Genitive (287-93); participial strategy impossible in verbless clauses (309)
 --- Distinct correlative strategy (349ff.): allows any element (including genitive) to be relativized, with interrogatives serving as WHICH-elements, remote demonstratives as THat-elements (350-51):
 [... WH-NP_i ... V-oo] [TH-NP_i ... V] ;
 however, this strategy not widely used whether in formal or colloquial language (Asher 25)
- 5) Special relative form of verb --- not exactly
 --- Adj.Ptcpl also used for "appositive adjectival clauses" (293ff.), i.e. the type "the fact that..."
- 6) Polypersonal verb --- 1 actant coded (Subject), suffixally
 --- Subj coding required in positive, nonmodal finite verbs, normally forbidden otherwise (Asher 184)
- 7) Infixing/suffixing alternation --- not applicable
 --- No Obj-coding on verbs; there are many compound verb forms, but always involving a "helping verb" which is (and behaves as) a real verb (317)
- 8) Definite article in genitive embeddings --- not applicable
 --- Genitive strategy: Possessor-gen Head (43), although "genitive" case-suffix doesn't always appear and other endings (not strictly "case" markers) are possible (42-46)
 --- No Def article; no exx. given involving Indef article on either N
- 9) Nonconcord of V with full-NP Subj --- NO
 --- Presence/absence of full-NP apparently not an issue (Asher 184-86)
- 10) Verbal abstract: VN or Inf? --- INFINITIVE
 --- Many types of nonfinite forms (70ff., 250ff.): Lehmann speaks of "infinitive" (common in EQUI contexts, 257ff.), Verbal.Ptcpl, Adj.Ptcpl,

Nominalization; exx. passim show that all of these take DObj in Acc (or Nom), never Gen, just as in main clauses; cf. Asher 184-86

11) Predicative particle? --- YES (maybe identical to "local" Postp)
 --- Copula optional, but two different constructions (171ff., Asher 49ff.):

Subj	Pred	Ø	(both NPs in Nominative)
Subj	Pred-aaka	COP	(nuance of temporariness, 175)

--- Copula omitted especially for present time (Asher 49)

--- This -aaka (Colloquial -aa) is an "adverbializing" suffix; suffixed to Noun, very like a case marker or Postp (146); uses, inter alia (139ff.):

Manner: speed-aaka = "fast"

Role: father-aaka = "(I'm speaking) as your father"

Result: ice-aaka = "(He made the water) into ice"

--- PredAdj follows same dual patterning: PredAdj has form of Noun if no overt Cop, form of Adv if there is an overt Cop (Asher 188)

--- Lehmann (146) analyzes -aaka as either "a bound postposition or a clitic"

12) Postpositional periphrastic (VN + Postp) --- a bit

--- Various Postps can take verbal-noun clauses as Obj (305ff.), but these all yield Adv clauses

--- There is a habitual construction of the form: VN + "BE" (using the verb uNTu "BE") [SH, p.c.]; but here VN is in Nominative

--- One example (305) cited as follows, using -aaka adverbializer (no further mention of this usage; a marked construction [SH, p.c.]):

he [talk-pres-Nmlzr-and laugh-pres-Nmlzr-and]-aaka BE
 "He is talking and laughing"

13) "DO" periphrastic (VN + "DO") --- in a sense

--- English verbs borrowed into Tamil are expressed by "Eng. Verb + DO" (Asher 207, 208), and even a few Tamil examples exist; Asher calls these "verbal compounds"; but the object of "DO" is (normally) a true noun, not a "VN" attached to some verbal paradigm [SH, p.c.]

--- Normally, "INF + DO" expresses Causative (Lehmann 219)

14) Adverbial clause = "and" + finite clause --- not applicable

--- "And" (-um) cannot directly coordinate finite clauses, only infinitives and Verbal.Ptcpls (appearing in compound verb-forms) (240-41)

--- Adv clauses expressed by a variety of nonfinite means: RC1 on "adverbial" HeadNoun (340ff.); "infinitive" clauses (261); Verbal.Ptcpl (272f.); verbal-noun clause + Postp (305ff.)

15) VN/Inf instead of finite main-clause form --- YES

--- Ordinary way of expressing (notional) sentence-conjunction is by using Verbal.Ptcpl (devoid of tense or pers/num/gender marking) for all but the last verb (265-66); usually but not always with identity of subject (deleted) (270)

--- Note that Verbal.Ptcpl does fulfill standard "Inf/VN" functions (73), as complement of Aux (193ff., 205ff.) and as Object complement of verb (274)

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- apparently NO

--- Dictionary mentions no relevant compounds [SH, p.c.]; no examples readily apparent in Asher's "Lexicon" section (244-53)

Tauya (New Guinea; Brahman stock-level family)
MacDonald 1990

0) Basic categorial information

--- Noun: marked suffixally for case and topic (85), but usually not for number (116) (though there is a "collective" ending *-?a*, 142ff.); cases listed 119ff.; case system Erg-Abs, but usage of Erg and Abs extremely flexible (315ff.)

--- Verb: Tauya has medial vs. final verbs; final verbs and most medial verbs marked for Subj (suffixally), Obj (prefixally) (171-72); Obj markers obligatory for humans, optional for animates, forbidden for inanimates (185)

--- Article: only indefinite article, either pre- or postposed (108); discussion of deictics (99ff.) includes no articular use

1) Conjugated adpositions --- barely (or irrelevant)

--- No formal mention of a category "adposition" under the lexical classes of Tauya (85); no discussion of such a category anywhere in the grammar

--- Though compare p. 4 ("Tauya includes postpositions"); and the term "postposition" recurs at least once, apparently used informally (108: the "indefinite postposition", referring to indefinite article)

--- There is a category "adverbial particle" (or adverbial modifier), which "take no inflection and function as sentential modifiers" (85); most of these are "uninflected independent forms" (279), but a few exx. seem to involve an accompanying preceding NP as "argument" (282-83); some of these particles are body-part nouns, hence rather like relational nouns in e.g. Mayan ("rib" = beside, "nose" = in front, etc.); since body parts take possessor prefix, these might conceivably be construed as "conjugated adpositions"; but the entire category seems in doubt

2) Word order

--- SOV unmarked (4), "fairly rigidly verb-final" (5)

--- Postpositions (4), if the category exists at all

--- N-Adj, usually (4); "Most qualifiers in Tauya follow the head noun" (105)

--- Gen-N (129-34); but "some genitives follow the governing noun" (4)

--- RCl-N, usually (105, 108ff.)

- 3) Relative clause linker --- take -na as linker [my analysis]
 --- RC1 ends with verb suffixed with -na; might be counted either as a linker, or as a morpheme defining a special RC1 form
 --- In favor of special-RC1-form analysis: the morpheme is presented as affixal (109)
 --- In favor of linker analysis: homophonous -na is the genitive marker (289-90), also marker of "complement clauses" (a type of "subordinate medial clause", 252), arguing for na as all-purpose subordinator [my own suggestion]
- 4) Relativization strategy/ies --- GAPPING, Gap/Copy
 --- Verb-final mood suffix is replaced by -na, and coref noun gaps (109, 289); order usually RC1-N, rarely the reverse (110); order N-RC1 primarily when RC1 consists of just the bare verb, and represents "a stable attribute of the head" (110)
 --- Relativization of obliques seems heavily dependent on recoverability of deleted noun; nouns marked adessive/allative may relativize, but not other local cases (293-94); a few Tauya speakers can relativize these, but only by using a Copying strategy, where resumptive pronoun allows preservation of embedded case-mark (294)
 --- No mention of relativizing Obj-of-postposition (category dubious)
 --- As for genitive relativization, may only relativize on inalienable possessors, because only here is a pronominal copy already present, cf. [8] below (Gap/Copy) (296)
 --- Also, gerundive nominalizations (see [10] below) can act as prenominal modifiers, like RC1 (110, 113-16)
- 5) Special relative form of verb --- no [my analysis]
 --- See [3] above for argument
- 6) Polypersonal verb --- 2 actants coded
 --- Obj prefixal (216-17), Subj suffixal (205-6); Obj markers basically identical to independent pronouns, but undergo special phonological processes as prefixes (216-17, cf. 92)
- 7) Infixing/suffixing alternation --- apparently NO
 --- Slot sequence (171) fixed: (Obj)-Stem-(Aux)-Subj-Mood
 --- For transitive verbs formed with transitivizing affix -fe-, human objects require special syntax, whereby fe alone takes verb-inflections and the lexical verb is isolated as a bare stem (180):
- | | | | |
|---------------|---------|---------------------------|----------------------|
| Human Obj: | ObjNoun | VerbStem | Obj- <u>fe</u> -Subj |
| Nonhuman Obj: | ObjNoun | [Verb- <u>fe</u>]-Subj ; | |
- the Obj-affix thus indeed shifts its position vis-a-vis the lexical verb; but fe would appear to count as the truly finite verb here (?)
- 8) Definite article in genitive embeddings --- not applicable
 --- No definite article; no relevant data re indef article
 --- Genitive syntagms (gen case markers -na, -pi) (129-34):

- Alienable: NounDept-na Head
Head PronDept-pi
(and NounDept can be doubled by PronDept)
- Inalienable: (NounDept-na) his/their-Head ;

3sg-possessed form (inalienable possession) involves zero prefix, but sometimes a special suppletive stem (130; cf. 296-97)

9) Nonconcord of V with full-NP Subj --- apparently NO

- Nouns seldom code plural (116); can use either a rare plural suffix (116), or the "collective" suffix, which "has a rather restricted distribution" (142) and is semantically heterogeneous (142-48)
- Only personal pronouns and verbs distinguish sg/pl (116)
- With human nouns (Subj or Obj), number indicated via choice of sg/pl pers/numb affixes on verb (118); nonhuman nouns take 3sg affixes on verb (194), hence notional category of number totally unexpressed
- Also, two verbal Aux markers have suppletive plural forms for all notionally plural nouns (regardless of animacy): one has suppletive form for plural Obj, the other for plural Subj (119, 194, 196)
- Exx. of conjoined Subj show verb agreement (135-38); no exx. of counted noun as Subj (116-18)

10) Verbal abstract: VN or Inf? --- apparently Inf

- Gerundive nominalizations with -mo, followed by case or topic marker (110-11); one ex. (111) shows normal Obj rection ("to twist rope")
- GerNom isn't fully nonfinite: must be marked with 1/2sg aorist Subj marker -e- (110, 320); however, this seems purely a frozen mark, playing no role in determining the semantic subject of the GerNom when embedded (320)
- GerNom may only take a 3pl Obj marker (115)

11) Predicative particle? --- sort of

- Apparently no copular verb, so equational sentences have NP predicates (162); Subj must take topic marker -ra
- Suffix -?a (162) used with such NP predicates (but also with verbal predicates, 208) to mark indicative mood; similar use of interrogative suffixes (163-64); equational syntagm (indicative) is thus:
Subj-ra Pred-?a

12) Postpositional periphrastic (VN + Postp) --- not applicable

- Apparently no postpositions, no copula

13) "DO" periphrastic (VN + "DO") --- NO

- Periphrastic use of transitivizer fe (see [7] above) might conceivably be relevant, but no direct indication that this should count as a verb at all (nor what this "verb" might mean)
- Compound verbs discussed (183-85), called "fairly productive", but no exx. with "do"; see p. 176 for lexeme meaning "do"

14) Adverbial clause = "and" + finite clause --- not applicable

- No mention of a clausal "and" conjunction; clause conjunction done by coordinate medial verbs (219); NP conjunction done either by

apposition or with comitative case (134ff.)

--- No discussion of adverbial clauses per se; apparently done either by locative/temporal relative clauses or by "subordinate medial verbs"

15) VN/Inf instead of finite main-clause form --- in a sense

--- Same-Subject coordinate medials (171), used in clause chaining, lack subject coding; but list of functions of coordinate medials (219-20) includes no prototypical VN/Inf usages

--- No indication that gerundive nominalizations can be used in place of "finite clauses"

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- no data

Turkish (Asia Minor; Turkic [Altaic])

Lewis 1967

[My thanks to Karl Zimmer for answers to several queries]

0) Basic categorial information: Agglutinating

--- Noun: suffixal coding of plural (25-26), case (28ff.), possessor (38ff.); no gender (25); order of morphemes is: N-Pl-Possessor-(n)-Case (40-41)

--- Verb: only subject coded (suffixally) (96ff.)

--- Article: only indirect article (bir = "one") (53-54)

1) Conjugated adpositions --- yes and no

--- "Primary" postpositions: take inflected independent pronouns (85ff.)

--- "Secondary" postpositions: really nouns, take possessive endings (89ff.)

2) Word order

--- SOV (240); not rigidly V-final (243-44)

--- Postpositions (85ff.)

--- Adj-N (239); rigid (242)

--- Gen-N (239); "definite izafet" (see [8] below) is reversible (242-43)

--- RCl-N (239) for participial RCl; also a literary, Persian-derived N-RCl construction (211ff.)

3) Relative clause linker --- ZERO (also invariant)

--- Participial RCl has zero linker

--- Persian-style RCl obligatorily uses conjunction ki (invariant): Head ki RCl (212)

4) Relativization strategy/ies --- Participle; Gap, Gap/Copy

--- Strategy: RCl-final V changes to participle, RCl immediately precedes Head (158ff.); coreferential NP gapped

--- Subj RCl uses uninflected participle (158, 260)

--- Oblique RCl (163, 261) uses "personal participle", with subject of RCl appearing as possessor suffixed to participle; a full-NP subject of RCl appears in ordinary "izafet" construction (see [8] below); coref NP deletes along with its case and/or (primary) postposition

--- Possessive RCl: use strategy for Subj RCl or for Oblique RCl, depending on whether the coreferential possessor-NP occurs as part of embedded Subj or embedded Oblique; possessor-NP is gapped, but possessive pronominal suffix survives (259ff.); examples of all types (260ff.):

[now speak-Ptcpl] man	"the man who is now speaking"
[write-Ptcpl-my] letter	"the letter which I wrote"
[father- <u>his</u> now speak-Ptcpl] <u>man</u>	"the man whose father is now speaking"
[picture- <u>his</u> -DAT look-Ptcpl-our] <u>artist</u>	"the artist at whose pictures we looked"

--- Also distinct, Persian-derived construction with finite RCl and Relative particle: Head ki RCl; alien to Turkish (211ff., 260)

5) Special relative form of verb --- not exactly

--- Personal participle can also be used as verbal abstract (165):

"(the fact of) my-doing" ;

and also forms the basis of various adverbial "gerund-equivalents" (184-85)

6) Polypersonal verb --- 1 actant coded (Subject), suffixally

7) Infixing/suffixing alternation --- NO

--- No object-coding on Verb; there are many compound tenses with "BE" (108), but "BE" acts as normal inflecting verb; "BE" immediately follows uninflecting main verb (sometimes written as suffix to main verb; only interrogative mark may intervene, 111), precluding any possibility of "infixing"

8) Definite article in genitive embeddings

--- Two types of possession ("izafet") constructions (41ff.):

- (i) Def: Possessor-gen Head-his ("of-the-expert his-report")
- (ii) Indef: Possessor Head-his ("Ankara its-city")

--- No definite article; indefinite article bir may appear on Head or Possessor or both (NB: a book-his = "a book of his") [KZ, p.c.]

9) Nonconcord of V with full-NP Subj

--- General rule (246): inanimate-pl Subj takes singular V; animate-pl takes plural V, but may also take singular to signify "a number of people acting as one"; also, inan-pl may take plural verb if Subj and V are widely separated

10) Verbal abstract: VN or Inf? --- INFINITIVE (usually)

--- Personal participle governs same (nonsubject) cases as finite verb (158), can be used to express verbal abstract (165)

--- Infinitive in -mek: government same as finite verb, per exx. pas-sim (167ff.); EQUI contexts

--- "Verbal noun" in -me normally governs Subj in Genitive, Obj in Accusative (as usual); if "Verbal-noun" is passivized, "Obj" appears (qua Subj) in Gen; less usually, can have Obj of active "Verbal-noun" in Gen ("the doing of this is easy") (170-71)

11) Predicative particle? --- NO

--- No particle; copula usually zero in present, nonzero otherwise:
Subj NomPred (Cop)

12) Postpositional periphrastic (VN + Postp) --- sort of

--- So-called "Present II" tense (111-12), used of "actions in progress", is formed by Infinitive in locative case, followed by conjugated verb "BE":

V-mek-te-BE = V-Inf-Loc-BE

13) "DO" periphrastic (VN + "DO") --- in a sense

--- Construction "Noun + DO" (et-) as a kind of compound verb (154), e.g. "do acceptance"; "Noun" is usually a foreign lexeme, often a foreign verb form (Arabic verbal noun, French past participle); only rarely Turkish, and then not deverbal [KZ, p.c.]

14) Adverbial clause = "and" + finite clause --- NO

--- Conjunction ve "and" (Arabic origin) little used in speech (206); also enclitic da "and then, also" (206ff.); neither is mentioned as being used for AdvClauses, which are rather made by various nonfinite "gerund" formations (174ff.)

15) VN/Inf instead of finite main-clause form --- weakly

--- Compound verb tenses built with "BE" need not repeat "BE"; e.g.

oda- m- da oturuyor(-du-m ve) gazeteyi okuyor- du- m
room-my-loc sitting and newspaper reading-WAS-1sg
"I was sitting in my room (and I was) reading the paper" (108)

--- If two verbs have identical suffixes, the first can instead take ending -ip (177-79):

kalk- t1- k git-ti-k ==> kalk-1p git-ti-k
rise-PAST-1pl go "We rose (and we) went"

--- NB: Neither of these involves Infinitive

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- NO [KZ, p.c.]

Tzutujil (Mesoamerica; Mayan)

Dayley 1985

[My thanks to Jon Dayley for answers to queries]

0) Basic categorial information

--- Noun: no cases; no gender; plural suffixes (139); possessive prefixes (= Ergative prefixes, 64, 141)

--- Verb: codes Subj and Obj person/number prefixally, on Ergative-Absolutive pattern (75); Obj marker (Absolutive) is termed "proclitic", but for no rigorous reason (see discussion, 138)

--- Article: both definite and indefinite articles (254-56)

1) Conjugated adpositions (relational nouns) --- YES

--- Relational nouns function as Preps, but take possessor-prefix Obj just like possessed nouns: "his-by" (= "by him") (152, 291-92); analyze as "hedged preposition"

--- Normally, possessor-prefix occurs on Relational noun even when full-NP Obj is present; but some RelNouns have special prefix-less short form used with full-NPs ("might be viewed as incipient prepositions", 152)

--- There are also four "true" Preps, central to the grammar; but no information on Pron object given (229, 291)

2) Word order (8)

--- VOS (VSO impossible) (304)

--- Prepositions (also relational nouns which precede their Obj) (229, 291-92)

--- N-Adj or Adj-N (194, 281)

--- N-Gen (though Possessor NP can be fronted, topicalization) (141, 281)

--- N-RC1 (281, 372)

3) Relative clause linker --- invariant or zero

--- Definite article ja(r) functions as relative marker (69, 231, 372); marker is optional (231); also serves as complementizer, cleftor (407)

4) Relativization strategy/ies

--- If relativized N is intransitive subject or transitive object, gap (373); if transitive subject, change to "focus antipassive" and gap (375)

--- Possessives always have the form (e.g.) "his-book John" (= "John's book") (141-42); RC1 on Possessor gaps full-NP ("John") (373), leaves pronoun ("Gap-Copy" type)

--- RC1 on Obj of relational noun treated like RC1 on Possessor (377)

--- If relativized N is instrument, change verb to Instrumental voice and gap (378)

--- If relativized N is Obj of a "true" locative Prep, gap Prep and Obj, but add enclitic resumptive particle wi7 just after verb (376); same technique possible (optionally) with the Relational noun "with" (377); same wi7 occurs when adverbial elements fronted (to preverbal position) in non-relative clauses (256)

- 5) Special relative form of verb --- NO
- 6) Polypersonal verb --- 2 actants coded on verb
 --- Language is ergative; order for transitive verbs is: Obj-Subj-V = Abs-Erg-V (75); Obj (Abs) marker is termed "proclitic", but this designation is in part arbitrary (138)
- 7) Infixing/suffixing alternation --- NO
 --- There are Aspect/Tense/Mood markers (79) and directional affixes (98-99) that are prefixed to verb, but none affects positioning of Subj, Obj; all these markers fall into a single slot-sequence
- 8) Definite article in genitive embeddings
 --- Definite and indefinite article may be used with pronominally possessed nouns ("the-my-sister") (255)
 --- N-Gen possession follows pattern "his-book John" (281)
 --- No explicit statement regarding placement of article(s) in genitive embedding (two relevant exx., pp. 285, 286); per Jon Dayley [p.c.], favored pattern is for definiteness to be marked once, either on Head or Dept, though double marking is possible
- 9) Nonconcord of V with full-NP Subj --- no mention, assume NO
 --- Grammar simply states that full NPs (Subj, Obj) are optional (298-99), since referenced in V
- 10) Verbal abstract: VN or Inf? --- BOTH
 --- Transitive verbs have "Active infinitive" in -ooj, and "passive infinitive" in -ik (incorporating a passive morpheme -j- in the V stem) (105)
 --- Infinitival phrases always lack Subj NP (380, 383, 395)
 --- "Active infinitive": takes notional Patient as Obj (381, 396), but can only be used when Patient is indefinite; this counts as INF
 --- "Passive infinitive": takes notional Patient as Poss prefix (383, 396); "despite the fact that the infinitive in the purpose clause is morphologically passive, the purpose clause ... has an overall active interpretation" (383); except for presence of passive mark -j- in stem, this is effectively a VERBAL NOUN (though grammar does not put it this way), because Obj shows up as genitive; e.g. (393):

 we-began [its-being-cut the-tree] = "We began to cut the tree"
 (i.e. "We began its-cutting/its-being-cut, the tree")
- 11) Predicative particle? --- NO
 --- Apparently either order (Subj NomPred, NomPred Subj) (302-3), but in any case NomPred (Noun or Adj) takes no particle
- 12) Prepositional periphrastic (Prep + VN) --- puzzling
 --- The syntagms given under [13] might seem apposite here, being Progressive constructions; but "be-in-the-act-of" appears closer to "DO" than to "BE", and "to" (a motion Prep) seems wrong for this type
- 13) "DO" periphrastic ("DO" + VN) --- YES, two uses
 --- (1a) Verb tajiin- ("be in the act of") functions as Progressive

auxiliary when followed by chi ("to") + INF (403, exx. p. 394)

--- (1b) Verbs b'anooj ("do"), majoon ("have begun" = "be in the act of") function as Progressive auxiliary when followed by INF (400, exx. pp. 393, 401); INF may sometimes be preceded by complementizer ja(r) ("for...to") (233, 391, 400)

--- (2) b'anooj + Spanish INF is "the primary way in which Spanish verbs are productively introduced into Tzutujil" (400), apparently without progressive nuance; syntax differs from (1b), in that notional object of Spanish INF appears as Obj of b'anooj (402): "they VERB you" is thus "you-they-DO VERBing"

--- See pp. 405-6 for complete list of auxiliary verbs

14) Adverbial clause = "and" + finite clause --- no mention, assume NO

--- Coordinate clauses may be joined by an "and" word, or be simply juxtaposed (360-61); time adverbial clauses may be formed by simple juxtaposition (thus "identical with concatenated coordinate sentences", 367), but no mention is made of linking with "and"

15) VN/Inf instead of finite main-clause form --- no mention, assume NO

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- no data

Wolof (West Africa; West Atlantic [Niger-Congo])

Njie 1982

[My thanks to Kevin Moore for advice and insight on many points]

0) Basic categorial information

--- Noun: no cases; noun classes (8 singular, 2 plural), each with its own class-marker (55ff.)

--- Verb: various series of clitic pronouns specialized as Subj and Obj person/number markers on Verb (101-2); these do not (NB) code noun class

--- Article: both indefinite and definite articles, each incorporating the class-marker (CLM) (60-61); Def Art postnominal, of form CLM-i/a; Indef Art prenominal, of form a-CLM

--- NB: Wolof is not a tone language (53)

1) Conjugated adpositions --- NO

--- Prepositions are followed by Series-XII pronouns (the "strong" series), equivalent to full NPs (101, 153); no explicit statement (though see 194), but inferrable from exx. (e.g. 234)

2) Word order

--- SVO (185)

--- Prepositions (233)

--- N-Adj (74); N and Adj joined by linker reflecting noun class (e.g. bu, for b-class Noun); but Adj appears to be a subclass of V [KM, p.c.;

cf. 225]

- N-Gen (73); N and Gen joined by linker i/u
- N-RC1 (86ff.); no explicit statement re order, but clear from exx.

3) Relative clause linker

--- Article (= class marker, 60) acts as RC1 linker (88ff.); confirmed per KM, p.c.

4) Relativization strategy --- GAPPING (some copying?)

- RC1 verb is normal finite verb; gapping of full NP (86ff. on RC1s)
- Normally, verb always takes subject clitic (see [6]); subject clitic present in Obj RC1, absent in Subj RC1 (92)
- No mention of oblique RC1s, except two exx. (92) demonstrating oblique relativization as ObjRel with voice changes on V (e.g. benefactive, cf. 198)
- Examples (not in grammar book; KM, p.c.) where Obj-of-Prep can be relativized, as follows: delete entire PrepPhr, but mark verb with Obj clitic ko "it", with (NB) no mark of voice change on V; construction puzzling; provisionally analyze as zero-marked voice changing

5) Special relative form of verb --- NO

6) Polypersonal verb --- 3 actants (max.) coded as clitics

- Person/number not coded affixally; but verb normally takes a subject clitic (101-2, 116-17), occurring in ten different series (conflating various aspectual/modal/emphatic nuances); the clitic cooccurs with possible full-NP subject
- Verb can also take object clitics (114, 118-19); two object clitics (IObj, DObj) can cooccur (193-94); a full-NP object is not doubled by an object clitic (118, 192-94)

7) Infixing/suffixing alternation --- YES

--- V may be preceded by Aux and/or Pre-Aux; Obj clitic is suffixed to Pre-Aux if it exists, else to Aux if it exists, else to V; and in several of the Subj-clitic series, Subj clitic follows a similar rule (168-69, summarizing 152-68); e.g. (here S and O indicate clitics):

Series II, VII, X:	V S O	Aux S O V
Series III:	S V O	S Aux O V
Other Series:	Pre-Aux S O (Aux) V	
	or S Pre-Aux O (Aux) V	

8) Definite article in genitive embeddings

- Article can cooccur with possessive pronouns ("my child the", 65, 71)
- Genitive embedding formed by: Head-i/u Gen
- With article, yields: Head-i/u Gen Art (73)
- Article may agree in class with either Head or Gen (per exx. 94-95), implying co-constituency with either Head or Gen; grammar says nothing about two cooccurring articles, but impossible per KM (p.c.)

9) Nonconcord of V with full-NP Subj --- YES (depends on analysis)

- Normally, full-NP subject is doubled by a Subj clitic on verb

(117); amounts to person/number "concord"

--- But in Subj-clitic Series I, III, V (in which [NB!] the Subj clitic is initial in the verb-complex), the subject clitic drops if full-NP subject occurs; this dropping is near-obligatory in series I and V, optional in III (155-56, 164-65); it's as if either a full-NP or a clitic but not both may occur at the very beginning of the verb-complex (my own suggestion)

--- Grammar (165) suggests reanalysis of Series I and V as (variant of) independent pronouns, not as clitics; pronoun (I, V) would then have status of full-NP; but on this analysis the verb in Series I,V would lack subject-clitic altogether (an ad-hoc stipulation)

--- NB: Series I is used as presentative ("le voilà qui crie", 143), Series V as emphatic ("c'est moi qui suis parti", 144); Series I is used with a "voici" morpheme, Series V incorporates a copula; so both of these seem conceptually cleft-like constructions, jibing with the pronoun's initial position in verb-complex; would seem to fit with suggestion that these are more like independent pronouns [all this my own speculation]

10) Verbal abstract: VN or Inf? --- INFINITIVE

--- "Infinitive" presented as a verb deprived of tense/aspect markers (112) and of person/number clitics (113); such infinitives usually preceded by complementizer a (113, 250), often incorporated into special subject-clitic pronoun series 10 ("X") and discussed in the grammar under Series-X pronouns (112-13, 172ff.):

bugga	ngeen-	a	lekka	"Do you want to eat?" (113)
want		you(X)	to eat	

--- Occurs after main verbs of feeling, will, obligation, or possibility (112); often this main verb (e.g. bugga) has been presented elsewhere as an Aux verb (121, 172ff.)

--- Exx. passim show this "infinitive" followed by Object-series pronouns (e.g. 172)

--- Also, instances that are presented as "EQUI" come up (e.g. 247); unclear if "infinitive" is intended here, too; have full finite verbs in many prototypical infinitive contexts (e.g. purpose clauses, 248)

11) Predicative particle? --- NO

--- Various syntagms: PredN SubjPron(IX) (112), others as well (see 224ff., 228), but never a particle

12) Prepositional periphrastic (Prep + VN) --- no mention, assume NO

13) "DO" periphrastic ("DO" + VN) --- no mention, assume NO

14) Adverbial clause = "and" + finite clause --- no mention, assume NO

--- Coordination, 276ff., including clause-level "and" conjunction (279-80); complement clauses (including Adv Clauses), 244ff.; no mention of requisite construction

15) VN/Inf instead of finite main-clause form --- no mention, assume NO

16) Word-initial change --- a bit

--- Nominalization of verbs can involve changes to initial consonant (39-40):

Vcd Stop	---	> Prenasalized stop
Vcls Stop	---	> (unchanged)
Fricative	---	> Vcls Stop (point of articulation may change)
Ø	---	> k

--- Initial change a major feature of closely related Fula

17) Extended use of kin terms ("Kin of Noun") --- apparently YES

--- Per dictionary (Fal et al. 1990): "child" and "fruit" are homonyms, but in different noun classes; 7th Muslim month = ndeyu-koor, "mother of koor", and 6th month = maamu-koor, "grandparent of koor", where koor is the 9th (!) month (Ramadan, cf. woor "to fast"; 8th month unrelated)

--- NB: Address-term (overheard by KM, p.c.) to a man whose job was planting trees, apparently a nonce coinage: baay garab "father of trees"; similar address-term baay sikkim "father of beard" (to a bearded man); these reveal the construction as productive and alive

--- Also: doom u reew "child of 'he nation" = "citizen"

Yagua (Amazon; isolate)

Payne & Payne 1990 (D. Payne 1986, D. Payne 1990)

0) Basic categorial information

--- Noun: no cases; many noun classifiers (445ff.), but not marked on N; number coded suffixally (451-52); Set I clitics as possessive "pre-fixes" (Set I) (361); no articles (349)

--- Verb: codes Subj with Set I proclitics (361), which can undergo major (i.e. affix-like) phonological fusion with V (364); clitic set distinguishes animate/inanimate gender, and (for animate) also person/number (403)

1) Conjugated adpositions --- YES (clitic-postposition) (378)

--- Patterns:	NP	Post	"Mamertu with" (363-64)
	Clit-Post		"him-with"
	Clit-Post NP		"him-with Mamertu" (rarest)

--- Identical to patterning with Gen-N (see [2] below)

--- These clitics objects belong to "Set I" (361) (used for verb subject, N possessor, object of postposition); do undergo phonological fusion with Postp (364)

--- Some postpositions always phonologically bound; some free when taking N object, bound when taking pronominal object (D. Payne 1990:124); the bound class should be reanalyzed as "hedged postposition", but will not take this as the dominant type; hence consider Yagua to have full-fledged postpositions

2) Word order

--- VSO (253, 259-60, 300)

--- Postpositions (378)

```

--- Class "Adj" barely exists (416); conforms to general N-Modifier
order (351)

```

--- Gen-N (348); Patterns:

Gen N "Thomas house"

Clit-N "his-house"

Clit-N Gen "his-house Thomas" (rarest)

--- N-RC1 (342), cf. general N-Modifier order (351)

3) Relative clause linker (342)

[illegible]

--- One (jirya'-tiy) is unmarked, invariant, equally usable with any sort of head noun

--- Others agree with head noun in e.g. animacy, number, etc.

4) Relativization strategy/ies (342-43)

--- Head + linker + normal finite clause; coreferential NP is usually omitted from RC1; presence/absence of resumptive pronoun depends on linker-type:

(a) If linker = jirya-tiy, then overwhelmingly do have resumptive pronoun (clitic); true for ANY role, including subject; clitics (Set I or II as appropriate) have the position and form they would have in a normal main clause (and indeed could co-occur with the NP in main clause, though rare)

(b) If linker = other, overwhelmingly do NOT have resumptive pronoun (again, for ANY role including subject)

--- Can relativize on most roles in relative clause; in particular:

(a) On Possessors (346, but no exx. given; a single ex. in D. Payne 1990:88-89); ex. involves copying (jirya-tiy), but presumably could involve gapping if clear in context [my inference]

(b) On "obliques", i.e. on Obj-of-Postp (346, also D. Payne 1990:88); exx. include two with copying (jirya-tiy), one with gapping (clear in context)

--- Also have headless RCl via nominalizations of verb (354, 357); distinct Agt-nominalization and Pat-nominalization suffixes

5) Special relative form of verb --- sometimes

```

--- NO for finite RCL with linker

```

--- YES for nominalizations

6) Polypersonal verb --- 1 actant (Subj) coded on V (Set-I clitic)

--- Subject clitics (Set I) prefixed to verb, or to Aux if there is one (see [7]); but no clitic if Subject is fronted to precede verb (361-62, 403)

Patterns: Subj Verb "Mamertu went" (fronted)
 Clit-Verb "he-went" (normal)
 Clit-Verb Subj "he-went Mamertu" (normal)

--- Object clitics (Set II) co-occur with full-NP object when object is "roughly definite"; must immediately precede Obj NP, and is phonologically enclitic to previous word; if no full-NP object, object clitic tends to be clause-final (364-67, 403)

--- Obj clitic only accidentally has Verb as its host word, hence not part of verbal personal marking

7) Infixing/suffixing alternation --- not really applicable

--- Obj not coded on V

--- There are 4 preverbal Aux elements ("could", malefactive, irrealis, negative) (362), which cannot take any normal verb suffixes or verb inflections (e.g. tense, location, imperfective) (414); hence not verbs in the usual sense

--- If an Aux element is present, Set-I clitics (Subj) are prefixed to Aux, not to main verb (362, 413-16); perhaps barely construable as positional mobility, though clitic always in initial position:

Clit - V Clit - Aux - V

--- But switch involves Subj, not Obj; and the alternation is not "infixing/suffixing" but "prefixing/prefixing"

8) Definite article in genitive embeddings --- not applicable

--- No definite articles (349) (though cf. [6] above for Set-II Object clitics with "roughly definite" objects)

9) Nonconcord of V with full-NP Subj (254) --- YES (inverse)

--- Patterns with Set-I clitics (D. Payne 1986:449, D. Payne 1990:241):

Subj V	Clit-V	Clit-V Subj	*Subj Clit-V	*V Subj
Gen N	Clit-N	Clit-N Gen	*Gen Clit-N	
N Postp	Clit-Postp	Clit-Postp N	*N Clit-Postp	

--- Thus, with verbs, pattern is REVERSE of Celtic/Hamito-Semitic: verb shows agreement (via clitic) if Subj follows V (the normal state of affairs), no agreement if Subj precedes V

--- Note similar pattern with Object: if Obj NP precedes verb, Set-II clitics do not appear (255)

10) Verbal abstract: VN or Inf? --- INFINITIVE

--- Ending -janu/-jada forms "action nominalizations", glossed as "INF" (352); but not purely the same as IE infinitive (333) (e.g. "sing-INF" = "song", "tell-INF" = "story") (352-53)

--- But also usable in prototypical infinitival contexts, e.g. EQUI (333), purpose clauses (337), e.g.

I-want sing-INF "I want to sing"

--- Infinitive can take subject pronoun (Set-I), construed as genitive ("his singing"); and examples show object pronouns in Set-II, i.e. object is normal "accusative" (337)

11) Predicative particle? --- NO

--- Order is PredNom Subj, with no copula and no particle (257, 371)

12) Postpositional periphrastic (VN + Postp) --- no mention, assume NO

- 13) "DO" periphrastic ("DO" + VN) --- no mention, assume NO
- 14) Adverbial clause = "and" + finite clause --- not applicable
 --- No word "and"; coordination usually by juxtaposition (294, 298);
 finite adverb clauses take various conjunction-like subordinating elements (339ff.)
- 15) INF instead of finite main-clause form --- no mention, assume NO
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- no data

Yimas (New Guinea; Lower Sepik family)
 Foley 1991

- 0) Basic categorial information: Polysynthetic (277ff.)
 --- Noun: 2 cases (unmarked and oblique, 90, 165); number (91); many noun classes, such that class/number inflection is conflated (92, 119ff.); all inflection suffixal; no article mentioned
 --- Verb: inflects (optionally! 227ff.) for Subj, DObj, IObj (person/number/class), usually with prefixes (93); Subj/DObj markers show complex split ergative patterning, including some 3-way marking (S, A, O) (195ff.), very complex ordering of affixes (202ff.)
- 1) Conjugated adpositions --- YES (106-9)
- 2) Word order
 --- Clause-level word order: Free (weak preference for V-final) (369-70)
 --- Postpositions (106-9)
 --- Two Adj constructions: (1) tight, Adj-N; (2) appositional, either order, Adj need not be adjacent to N (93-94, 182-84)
 --- Only a handful of true Adjectives (93-94); otherwise have Adjectival verbs (94ff.) and Adjectival nouns (98ff.)
 --- Two Gen constructions: (1) tight, Gen-N; (2) appositional, either order, Gen need not be adjacent to N (176ff., 180-84) (see also [8] below)
 --- RCl: either order, RCl need not be adjacent to Head N (420)
- 3) Relative clause linker --- ZERO
 --- m- marker on V (see [4]) is not constrained to occur at RCl boundary, hence not "linker"
- 4) Relativization strategy/ies --- GAPPING (?)
 --- In finite RCl, internal order of constituents unconstrained
 --- V usually takes an m- prefix (= near-distal deictic) (413), but not with Neg (419) or with adjectival RClS (428)
 --- V takes an added class/number suffix agreeing with head, basically

identical to adjectival concord suffixes (413); special suffix -nak used only for Obj RC1 of singular human head N (415), functionally rather like an Indo-European "relative pronoun" in that it marks role of coref N in RC1 ("whom" vs. "who")

--- Coreferential N and its cross-referential Subj/Obj affix on V must both be absent from RC1 (413ff.); coref N signalled only by the added class/number suffix on the V, which is not a clause-internal pronominal copy but an agreement marker with head N; hence "Gapping"

--- Can relativize on any core argument, also locatives and temporals (405, 418); no mention (?) of genitival or postpositional relativization

--- Also nonfinite relativization (Agent nominalization) (403ff.)

5) Special relative form of verb --- YES (m-)

--- But m- not obligatory, sometimes forbidden; slight difference of nuance for RC1 with and without m- (428-29)

6) Polypersonal verb --- 3 actants coded (193ff., esp. tables 217-18)

--- Usually prefixes, except 3-person Dative is suffixal

--- Complex rules of prefix ordering; if 3rd- and 1/2nd-person actants cooccur, always have order "3Pers-1/2Pers-Verb" regardless of function

7) Infixing/suffixing alternation --- YES (suffixing/prefixing)

--- If verb takes NEG or POT(ential) modality prefixes, these usurp prefix-slot of leftmost Subj/Obj prefix (with transitive verbs); that actant resurfaces as a class/number suffix (251ff.), basically identical to adjectival concord suffixes; suffix does not code person, but since the leftmost preverbal Subj/Obj prefix is usually 3person anyway, not much ambiguity results

--- Relative-clause m- marker has similar reordering effect (see [4]), now even with intransitive verbs

8) Definite article in genitive embeddings --- not applicable

--- No articles; genitive syntagms are (180ff.) (see [2] above):

(a) Rigid: Possessor-na HeadN

(b) Appositional: Possessor-na-ClM HeadN
(where ClM is class-marker of HeadN)

9) Nonconcord of V with full-NP Subj --- optional

--- All verbal pronominal affixes optional (227ff.), notably if cooccurring with full-NP which is new information (232ff.)

10) Verbal abstract: VN or Inf? --- INFINITIVE

--- Non-finite nominalizations (377ff.) clearly coded as such (suffix -ru), lack tense (377) and Subj/Obj affixes (392); subject appears as possessive (392-93); no explicit statement about Obj, but exx. passim show no change of Obj to possessive

11) Predicative particle? --- NO

--- Use copula, no particle (225-27); AdjPred usually a verb

12) Postpositional periphrastic (VN + Postp) --- no mention, assume NO

- 13) "DO" periphrastic ("DO" + VN) --- sort of
 --- Serial-verb construction can have "DO" (actually "do/feel/become") as first verb comprising a serial-verb complex, with two specialized nuances (e.g. "do V extra-thoroughly") (334-36)
- 14) Adverbial clause = "and" + finite clause --- no mention, assume NO
 --- The only "and"-type conjunction is "and then" (116); discussion of coordination (449ff.) does not indicate any adverbial use
- 15) VN/Inf instead of finite main-clause form --- weakly
 --- Nonfinal verbs in clause-chaining are "Dependent verbs" marked with ending -mpi (446), devoid of person/number marking; same morpheme can also be used in serial-verb combinations (322, 326); the form is distinct from the infinitive-type nominalization of [10] above
- 16) Word-initial change --- no mention, assume NO
- 17) Extended use of kin terms ("Kin of Noun") --- no data

Yoruba (West Africa; Kwa [Niger-Congo])

Rowlands 1969; also Awobuluyi 1978, Bamgbose 1966, 1974, Abraham 1958; see also Welmers 1973 (passim)

- 0) Basic categorial information: Isolating
 --- Noun: no gender, case, articles; plural optionally indicated by adding a free pronominal "they" before Noun (1969:40)
 --- Verb: no arguments coded on verb; distinct series of free pronouns (Subj, Obj, Possessor) (1974:10-11); Subj pronoun does not normally cooccur with full-NP Subj (per exx. passim), but subject NP normally is followed by a high tone which can be taken as a manifestation of the 3sg Subj pronoun and occurs wherever that pronoun could occur (Welmers 1973:381-82, cf. 1969:34); no sign that pronouns are bound more tightly to Verb than are full-NP arguments
 --- Article: none mentioned; there is a "previous-mention" Demonstrative noun-qualifier naà (1969:197, 1974:14), sometimes translatable "the" (1978:35); will not reanalyze as "hedged article"
 --- Caveat: Different grammar books present central aspects of Yoruba in many different lights, not easy to reconcile
- 1) Conjugated adpositions --- NO (or not applicable)
 --- "Preposition" a marginal category, status disputed; Awobuluyi admits 7 Preps (1978:97, cf. 76), yet one of these (ex. 57) undergoes the kind of reduplication nominalization seen for verbs (cf. [10]); Bamgbose (1974:19, 22, 33) presents these as verbs; Rowlands (1969:21) says ní "to, in, at" is the only real Prep, though others have Prep-like uses (82-84); per Abraham, "certain words now used as prepositions etc. were originally verbs" (xiv); Welmers presents several of these as serial verbs (1973:376), accepts ní as Prep (453)
 --- Will here take only ní as Prep; but (1969:21) ní cannot be

followed by an unemphatic pronoun, unclear if it can take pronominal Obj at all; the other "Prep"s, like any verb, take pronominal Obj as separate word

--- Many prepositional notions expressed by body-part terms, directional adverbs, nouns, etc. (1969:139ff.)

2) Word order

--- SVO (1978:111)

--- Prepositions (insofar as the category exists) (1978:97)

--- N-Adj (1969:125, 1974:16)

--- N-Gen (1969:44, 1974:16)

--- N-RC1 (1969:87, 1974:16)

3) Relative clause linker --- INVARIANT tí (1969:87)

4) Relativization strategy/ies

--- Syntagm: HeadN tí RC1 (1974:15, 1978:94ff.)

--- In Subj RC1, if HeadN is 3rd-person (sing or pl), coref NP in RC1 replaced by singular resumptive pronoun ó "s/he/it" (1978:94, 1969:87-88); if HeadN is a 1/2-pers emphatic pronominal, RC1 has corresponding person/number Subj pronoun

--- In Obj RC1 and Prep RC1s, coref NP gaps (1978:95); if Preps are taken as Verbs, this collapses into a single rule; recall that ní (the only true "Prep") could not take an unemphatic pronoun anyway (1969:21)

--- In genitival RC1, coref NP leaves pronoun copy (1978:96, ex. 1966:119)

--- NB: pattern of RC1s is exactly like pattern of cleft sentences, except that clefts use particle ní instead of tí (1969:90)

5) Special relative form of verb --- NO

6) Polypersonal verb --- no actants coded

--- Special tone rules may sometimes apply to Subj and Obj pronouns (1974:10-11); but not an indication of combination with verb (?)

7) Infixing/suffixing alternation --- not applicable

8) Definite article in genitive embeddings --- not applicable

--- No article; syntagm is HeadN - Gen;

final vowel of HeadN is lengthened on a mid-tone before consonant-initial noun or before certain pronouns (1974:12-13, 1966:110, 1969:45)

--- Alternate syntagm (emphatic): HeadN tí Gen (1969:45); note that tone of tí differs from RC1 marker

9) Nonconcord of V with full-NP Subj --- not applicable

--- But in Subj RC1, note singular resumptive pronoun even if HeadN is plural

10) Verbal abstract: VN or Inf? --- apparently Inf

--- Several deverbal noun formations (Agent, Patient, Action nominalizations), marked by various prefixes (1978:87ff., 1969:182ff.); all are stated to nominalize full VPs as such (1978:89); only some formations have exx. displaying transparent accusative rection (i.e., use of Obj pronouns, not possessive pronouns), but none appears to show clear

genitival rection (i.e., no characteristic vowel lengthening [see 8])

--- Will examine three such formations (A, B, C):

[A]: Reduplication in C₁- (1978:88) functioning as verbal abstract, very productive (1969:189); uses:

(i) Preceding the identical finite verb, to emphasize verbal action (as if "sing-Inf he sang") (1969:189)

(ii) In constructions like "It is difficult to do":

ó sòro ní s₁ise it is-hard at/in do(Redup),

often with reduplication prefix reduced (1978:57, 118-19; cf. 1969:191)

(iii) As cognate object following a verb (1978:51, 124)

[B]: Productive use of preverbal nominalizer àti, equivalent to English infinitive (1969:188-89), accusative rection (exx. 188); spelled as separate word; negative equivalent uses prefix àì- (1969:192-93)

(with Subj expressible as Possessor)

[C]: In EQUI contexts, certain main verbs occur with a lengthened vowel (1969:66, 191; slightly differently 1966:76-77, 1974:20-21, with added vowel specified as high-tone); lengthening "is best regarded as really a prefix of the following verb", which may thus be considered an "infinitive" (1969:67, cf. Welmers 1973:358); accusative rection (ex. 67); perhaps analyzable as reduced form of (A) above (1969:191)

--- Cf. Abraham's dense and puzzling account of "verbal nouns" (1958:xxvi-xxvii)

11) Predicative particle? --- NO

--- Basic syntagm: Subj Cop Pred, with a variety of copula verbs (1969:152, 158)

12) Prepositional periphrastic (Prep + VN) --- apparently NO

--- Isolated ex. glossed: "they are in standing-position" (1969:154); here "standing-position" is (nonreduplicated) nominalization of "stand"; but no indication that this represents a productive pattern

--- Progressive marker is verb prefix n̄- (1969:60); Welmers's (1973:313-14) suggestion of some ultimate kinship with Prep nì "in" is synchronically weak (allomorphy quite different)

13) "DO" periphrastic ("DO" + VN) --- YES (with a twist)

--- Verb se "do, make" is not singled out as preverb or postverb (1974:19ff.), i.e. not an Aux

--- But there is a special combination of se + negative infinitive in àì- (see [10] above), fusing to yield sàì + VERB; when cooccurring with another negative, sense is a strong affirmative (1969:192-93, cf. 1958:607):

Neg...sàì + VERB = "Neg do Neg-VERB" = "to not fail to Verb"

--- Additionally, certain phonesthemes form composite verbs with se (1969:122, 154-55), though usually rì "get, be" is used instead

14) Adverbial clause = "and" + finite clause --- NO or not applicable

--- No true clause-level "and" conjunction (1978:105, cf. 1969:203); but have an Aux verb sì, not really meaning "and" but "(be) also, moreover" (1969:203, 1974:19, Welmers 1973:377), with no apparent Adv uses

--- Adv clauses formed variously, e.g. with particles (e.g. 1974:33)

15) VN/Inf instead of finite main-clause form --- apparently NO

--- No main-clause use evident for formations in [10]
 --- Much serial-verb formation (1969:132ff.); but this is not clearly "nonfinite"

16) Word-initial change --- no mention, assume NO

17) Extended use of kin terms ("Kin of Noun") --- YES

--- Checked large dictionary (Abraham 1958) s.vv. qmqn "child", baba "father", iya "mother", and cross-references; numerous lexicalized combinations listed, though many don't fit the requisite semantics precisely; "child" can mean "pip of fruit, kernel of nut" (518); also "child" + placename = someone born in a place (518)

--- Further: "child padlock" = key of padlock (23), "child work" = apprentice (321), "child outside" = hooligan (325), "child mortar" = pestle (451), "child war" = soldier (456), "child death" = orphan (467), "child presence" = body-servant (504); "father (of) deceased" = deceased's executor (92; unsure of my analysis)

--- Presumably all are genitive combinations, though the characteristic lengthened vowel of genitive (see [8] above) is not shown in dictionary (per standard orthography, 1969:45)