Constructional Morphology:
The Georgian Version

by

Olga I Gurevich

B.A. (University of Virginia) 2000
M.A. (University of California, Berkeley) 2002

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, BERKELEY

Committee in charge:
Professor Eve E. Sweetser, Co-Chair
Professor James P. Blevins, Co-Chair
Professor Sharon Inkelas
Professor Johanna Nichols

Spring 2006

The dissertation of Olga I Gurevich is approved:

Co-Chair Date

Co-Chair Date

Date

Date

University of California, Berkeley

Spring 2006

Constructional Morphology:
The Georgian Version

Abstract

Constructional Morphology:
The Georgian Version

by

Olga I Gurevich
Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Eve E. Sweetser, Co-Chair,
Professor James P. Blevins, Co-Chair

Linguistic theories can be distinguished based on how they represent the construction of linguistic structures. In "bottom-up" models, meaning is carried by small linguistic units, from which the meaning of larger structures is derived. By contrast, in "top-down" models the smallest units of form need not be individually meaningful; larger structures may determine their overall meaning and the selection of their parts. Many recent developments in psycholinguistics provide empirical support for the latter view.

This study combines intuitions from Construction Grammar and Word-and-Paradigm morphology to develop the framework of Constructional Morphology. The proposed framework provides mechanisms for describing the full range of regular, sub-regular and irregular patterns in languages with rich morphology and complex morphosyntax.

The thesis argues that morphological and morphosyntactic patterns should be described using generalized form-meaning pairings (*constructions*), which include semantic, syntactic, and morphological information in the same statements. This top-down approach also resolves some long-standing issues in computational morphology.

The theoretical framework is illustrated through an analysis of Georgian morphosyntax with a particular focus on *version*, originally a morphosyntactic marker of participant affectedness or salience. Version represents a case of mismatch between form and function: the same morphological resources can mark participant affectedness in some constructions and unrelated categories in other contexts, such as voice, tense, and conjugation

class. The syntactic function of version markers is in some contexts akin to an applicative, elevating an affected participant to a syntactic core argument, while in other instances they make no syntactic contribution.

The theoretical framework, developed to capture the recurrent patterns of Georgian morphosyntax, is also applicable to general morphosyntactic description. An examination of version-like phenomena in several other languages reveals that their description also depends on the larger constructions of the particular language and should therefore be done in the same "top-down" approach.

The thesis concludes by exploring the consequences of Georgian-type patterns for computational linguistics, which has traditionally assumed straight compositionality. A computational model is proposed for parsing and generating Georgian verbal inflections based on example paradigms and constructions at various degrees of generality.

Professor Eve E. Sweetser
Dissertation Committee Co-Chair

Professor James P. Blevins
Dissertation Committee Co-Chair

To the memory of my grandmother, Yudif Gurevich (1912-2006)

Contents

List of Figures

List of Tables

Acknowledgments

Dissertations are not written in a vacuum, and this dissertation owes a great deal to the help and support of many people. My advisors, Eve Sweetser and Jim Blevins, have tirelessly guided and encouraged me through the whole long process. I would like to thank Eve for inspiring me to become a cognitive linguist and to look beyond form alternations to the larger context in which language is embedded. I am grateful to Jim for introducing clarity to the complexity of the data, and for teaching me to think critically about linguistic patterns and linguistic theories. Thanks to the other members of my committee: Sharon Inkelas for picking holes in my argumentation and forcing me to think through the details, and Johanna Nichols for encouraging me to work on the Caucasus and for making me interested in typology.

Other members of the Berkeley faculty have been immensely helpful in my progress through the program. Thanks to Andreas Kathol for helping me through morphosyntactic hurdles at the earlier stages; to George Lakoff for making me come to Berkeley in the first place; to George Lakoff and Jerry Feldman for many discussions of embodiment and cognition; to Charles Fillmore and Paul Kay for coming up with Construction Grammar; to Carla Hudson Kam for discussing language acquisition with me; to Lynn Nichols for help with qualifying papers and syntax; to Andrew Garrett and Gary Holland as graduate advisors for making my choices easier; and to the rest of the Linguistics faculty for making the program as diverse and interesting as it is.

I cannot fully express my gratitude to Shorena Kurtsikidze and Vakhtang Chikovani, who taught me Georgian, encouraged me to study it further, and provided most of the data in this dissertation; without them, my path through Berkeley would not have been nearly as exciting (and challenging). დიდი მადლობა, შორენა და ვახტანგ!

This dissertation could not have happened without the foundational English-language work on Georgian by Howard Aronson, Alice Harris, Winfried Boeder, and Kevin Tuite. I particularly want to thank Alice Harris for providing extremely useful, detailed comments on my papers and presentations; Winfried Boeder for letting me use his personal library; Winfried and Doris Boeder for inviting me, a stranger, to their home in Oldenburg; and Kevin Tuite for encouragement and comments on papers. Thanks to Nino Amiridze for insightful suggestions on Georgian data, and to Vano Nasidze for lending me his native-speaker intuitions. Thanks to the Linguistics department at the Max Planck Institute for

Evolutionary Anthropology, and particularly to Bernard Comrie and Martin Haspelmath, for letting me spend a very productive semester in Leipzig. Thanks especially to Bernard Comrie, Martin Haspelmath, Orin Gensler, and Jeff Good for discussing the first stages of my research on Georgian version and for pointing me to relevant typological literature. Special thanks to Jeff Good for help with LaTeX. Thanks to Greg Anderson for making me realize 'version' was ubiquitous, and to the audiences at various conferences where I presented my work for crucial questions and suggestions.

Thanks to the Berkeley Language Center, and particularly to Claire Kramsch, Mark Kaiser, Lisa Little, and my fellow BLC fellows, for the financial, logistical, and intellectual support in the development of the Georgian computational model and learner's reference. I am grateful to Lauri Karttunen for introducing me to finite-state morphology and for providing software updates.

Thanks to the folks at the Berkeley Neural Theory of Language group, especially Nancy Chang, Ben Bergen, Ellen Dodge, and Shweta Narayan, for listening to me blabber about morphology, and for help with constructional approaches. My classmates in the Berkeley Linguistics department created a friendly, supportive environment that made it all worth it. The wonderful staff of the Linguistics department: Belén Flores, Paula Floro, Esther Weiss, and Ron Sprouse, have provided invaluable logistical support and advice throughout my stay at Berkeley.

Finally, thanks to my family for being supportive and patient throughout my graduate career, even though they may have wondered at times what I was doing working on strange languages. Thanks to my husband Aaron Siegel for being a wall of strength when I needed him most, as well as for being willing to read through texts he didn't understand and for much technical help.

Abbreviations in the Georgian Data

1, 2, 3	person	OPT	optative
AOR	aorist	PASS	passive
CAUS	causative	PRESSUBJ	present subjunctive
COND	conditional	PL	plural
DAT	dative	PERF	perfect
DO	direct object	PLUPERF	pluperfect
EM	extension marker	PERS	person
ERG	ergative	PRES	present
FUT	future	PRON	pronominal agreement
FUTSUBJ	future subjunctive	PRV	pre-radical vowel
GEN	genitive	PV	preverb
IO	indirect object	RV	root vowel
IMPERF	imperfect	SG	singular
INTR	intransitive	SCR	screeve
NAR	narrative	SUBJ	subject
NOM	nominative	TAM	tense/aspect/mood
NUM	number	TS	thematic suffix
OBJ	object	VER	version

Abbreviations in Other Data

ABL	ablative	MOD	modal
ACC	accusative	NEG	negation
AFF	affective	NEUT	neuter
AUX	auxiliary	NPAST	non-past
AV	auxiliary verb	OBL	oblique
CLOC	cislocative	PROG	progressive
CON	converb	PST	past
CV	converb	REFL	reflexive
FEM	feminine	SS	same subject
FOC	focus	VERS	version
IRR	irrealis	VC	voice
LV	lexical verb	Q	question
MASC	masculine		

Other Abbreviations

CxG	Construction Grammar
FSA	Finite-State Automaton
FSN	Finite-State Network
FST	Finite-State Transducer
IA	Item and Arrangement
IP	Item and Process
HPSG	Head-Driven Phrase Structure Grammar
LFG	Lexical Functional Grammar
WP	Word and Paradigm
XFST	Xerox finite state transducers

Chapter 1

Introduction

1.1 Overview of the Study

This study is about the relationship between form and meaning in morphology. The major question addressed here is the nature of correspondence between 'units of form' and 'units of meaning' in languages with complex morphology and morphology-syntax interfaces. As part of this broader question, the study questions whether the form-meaning relationship can and should be described as one-to-one (as is done in many traditional approaches using morphemes), or whether the form-meaning relationship is more complex and requires more flexible methods of description.

My major point is that morphological descriptions can in fact be simpler and more natural when both form and meaning are allowed in. I argue against strict separation into levels such as semantics, syntax, morphology, etc. Rather, I hope to demonstrate that semantic and even discourse / pragmatic restrictions can help in descriptions of morphologically complex languages.

More specifically, I will show that:

- Meaning construction at the sub-word level *can* be just as complex and non-compositional as in syntax.
- It often *is* complex and non-compositional.
- In fact, this might be a more common state of things than the full compositional transparency.

Based on these main points above, I will argue for the use of form-meaning pairings (*constructions*) in describing morphological systems and their connection to syntax and

semantics. I will suggest a holistic, top-down method of description which includes semantic, syntactic, and morphological information in the same statements rather than a bottom-up method which builds up meanings compositionally from smallest units (i.e. morphemes).

This mode of description is supported by emerging evidence from psycholinguistics and language acquisition. I will argue that it also helps solve some long-standing issues in computational modeling of morphologically complex languages.

The theoretical framework underpinning much of the analysis in this study is Construction Grammar (CxG), broadly construed. The fundamental insight of CxG is that the smallest unit of linguistic description is a form-meaning pair called *construction* which could be as small as a morpheme and as large as a phrase. A construction may involve a combination of several smaller elements, but the meaning of the combination is not necessarily linearly composed (or predictable) from the meanings of its parts. Rather, the combination itself may carry aspects of meaning which add to or override the meanings of the parts. As a general account of linguistic patterns, this model must be able to account for cases that are taken to support a compositional, morpheme-based model as well the more complex form-meaning relations. I will suggest that the truly 'compositional' patterns are easily accommodated within the same descriptive approach, but the patterns themselves may not be the default in language. Rather, they may emerge as a useful strategy during language acquisition (cf. Blevins 2002 for a similar argument).

Construction Grammar has so far been applied to problems in syntax and semantics but not to morphology, and in particular, not to languages with rich inflectional morphologies. In principle, however, there is no reason why it could not be used to describe such languages. This thesis is thus a case study in Construction Grammar, expanding the framework while solving a traditional problem in grammatical analysis. Chapter 2 includes a more detailed description of the framework, its background, and morphology-specific adaptations. The general principles of CxG will be further supported by psycholinguistic evidence from several languages.

The theoretical framework proposed in this dissertation will illustrated through a description and analysis of Georgian morphosyntax. In particular, I will focus on on *version*, originally a morphosyntactic marker of participant affectedness or salience. Version represents a clear case of mismatch between form and function: the same morphological resources can mark participant affectedness in some constructions and unrelated categories in other contexts, such as voice, tense, and conjugation class. Syntactically, version markers

in some contexts function similarly to an applicative, elevating an affected participant to a syntactic core argument, while in other instances they make no syntactic contribution. More generally, version interacts with the systems of verbal person-marking and indirect object coding, as well as voice and valence alternations. I will analyze the morphosyntactic properties of version and argue that certain syntactic alternations are driven by verbal semantics and are best described constructionally. I will also analyze the semantics of version and argue that its different functions can be analyzed as a family of related constructions.

Chapter 3 is a general description of Georgian verbal morphology. In particular, it demonstrates several properties of the system that make it particularly amenable to a constructional analysis. These include the distribution of agreement markers, preverbs, thematic suffixes, stem formation, and the morphosyntax of various TAM paradigms.

Chapters 4 and 5 examine the function and distribution of *pre-radical vowels* in Georgian, which provide the formal means of encoding version as well as a set of other, seemingly unrelated contrasts. Chapter 4 focuses on the semantic and syntactic function of the core version sub-system and argues that the function of pre-radical vowels crucially depends on the larger morphosyntactic constructions in Georgian, particularly the systems of voice and valence encoding. Chapter 5 covers the uses of pre-radical vowels that are not semantically motivated but rather fulfill strictly morphological needs. Once again, the distribution and function of these vowels can only be described in the context of other construction types of Georgian. This chapter also traces some synchronic and historical connections between the different uses of pre-radical vowels and suggests that the two sets of connections are not isomorphic.

Chapter 6 exemplifies a further extension of the constructional view to the field of computational language modeling. It suggests that the standing assumptions in computational linguistics are largely influenced by the morpheme-based approaches to language analysis and suggests ways in which a constructional view could benefit computational modeling. The chapter describes a computational model for parsing and generating a subset of Georgian verbal paradigms and discusses the relevance of this approach to solving some long-standing computational issues in modeling morphologically complex languages.

Finally, Chapter 7 sums up the findings of the first six chapters and suggests possible extensions of the Constructional Morphology framework to similar structural phenomena and to morphological patterns more generally.

Below are some of the major implications and issues of the work presented in this

dissertation.

- This thesis provides a comprehensive description of Georgian version and parts of Georgian verbal morphology that have not been fully described in the Western linguistic literature.
- It proposes a modern, psycholinguistically plausible theoretical framework that relies on many older, traditional intuitions about linguistic structure. This model is capable of describing regular, irregular, and sub-regular morphosyntactic patterns in the same representational format.
- The proposed framework is used to analyze a traditional problem in language description that has nevertheless presented much difficulty for analysts in the past.
- What, if anything, is a morpheme and how useful is the concept?
- Should we get rid of compositionality? At the very least, it must be redefined to include combinations of templates (constructions), not necessarily linearly, and only in pre-specified contexts. In other words, it is a much more limited notion, one that is not rule-based but emerges as an abstraction over many specific instances of constructional combinations.
- A field linguist should be able to detect constructional phenomena in her language. The study provides some guidelines for detecting and analyzing such phenomena.
- The current approaches to computational morphology mostly operate with morphemes. This study suggests some ways in which constructions can be incorporated into computational tools and provide a computational model for a subset of the Georgian verbal morphology.

After a note on terminology and transliteration, the rest of this chapter provides a brief introduction to the main focus of this thesis, Georgian version, and examines some functionally similar cross-linguistic phenomena.

1.2 Terminology

Georgian belongs to the Kartvelian language family, which also includes Svan, Mingrelian, and Laz. This family is also sometimes termed South Caucasian, but the term Kartvelian will be used here[1]

[1] Amiridze (2006:1) comments that the use of the term South Caucasian may suggest a genetic relationship between Kartvelian languages and those of the North-West or East Caucasian language families, which at

There has been an overabundance of terminology for the different elements of Georgian morphosyntax, from the Georgian as well as Western linguistic traditions. I chose terms that seemed to be most neutral, presupposed the least amount of theoretical bias, and best reflected the functionality of the category described.

The set of vowels used to encode version are the central focus of this thesis. These vowels have been variously called version vowels, character vowels, theme vowels, pre-stem, and pre-radical vowels. I will use Aronson's (1990) term *pre-radical vowels* and occasionally, when describing instances of real versional meaning, the term *version vowels*.

For case names, I use the traditional terms: Nominative, Narrative, Dative, and Genitive. The Narrative has also been called Active or Ergative, underlying a theoretical debate to be discussed in Chapter 3.

The inflectional categories of verbs (corresponding to tense paradigms in other languages) are traditionally called *screeves*, from the Georgian *mc'k'rivi* 'row'. The term screeve implies a combination of tense, aspect, and mood properties, and is thus more appropriate for the description of Georgian forms, and was suggested for use in English by the great Georgian linguist Akaki Shanidze (as quoted in (Aronson 1991)).

The screeve names have been chosen mostly following (Harris 1981) and are listed in table 1.1, along with a partial list of alternative names for each.

SCREEVE NAME	ALTERNATIVE NAMES
Present	Present Indicative
Imperfect	Past Continuous, Past Imperfective
Present Subjunctive	Conjunctive Present, 1st Subjunctive
Future	Future Indicative
Conditional	Past Frequentative, Future Imperfective
Future Subjunctive	Conjunctive Future
Aorist	Aorist Indicative
Optative	Aorist Subjunctive, 2nd Subjunctive
Perfect	1st Resultant, 1st Evidential
Pluperfect	2nd Resultant, 2nd Evidential

Table 1.1: Screeve Names

present has not been established.

The terms Agent, Patient, and Goal are intended to designate basic participants in a transitive scene, or Manipulative Activity Scene (Slobin 1985), and can also be conceived as general clusters of properties that Dowty (1991) has called *proto-roles*. As much as possible, I try to avoid any other theoretical notions associated with these terms. The more detailed explanations of terminology will be given when the terms are used.

1.3 Transliteration

Georgian has a unique alphabet that matches its phonemic inventory remarkably well. A commonly-used transcription in that used by (Harris 1981:xvii), which uses under- and over-letter diacritic marks to indicate ejectives and several fricatives. One of the goals of data collection in this study was to make it maximally easy for archiving. I therefore used combinations of plain-text characters wherever possible. So while the transcription adopted here is largely the same as the frequent one, there are minor differences, as indicated in Table 1.2.

Transliteration in Russian examples is done according to the US academic standard, and for data from other languages it follows the original sources.

1.4 Brief Introduction to Georgian Version

This section aims to demonstrate that Georgian version is a truly constructional phenomenon, for several reasons. Here, I provide a cursory overview of the system; see Chapters 4 and 5 for a more detailed analysis.

The term "version" originates from the Georgian grammatical tradition; it is a translation of *kceva* (lit. 'change'). In Georgian, the category of version interacts with voice and Tense-Aspect-Mood (TAM) properties (Tschenkéli 1958, Boeder 1968, Hewitt 1995b). The Kartvelianist grammatical literature describes version as one of the inflectional categories of a verb, along with TAM, person, number, and others.

Formally, version is marked by a set of "pre-radical" ("pre-stem") vowels, the historical origins of which are unclear. These so-called *version vowels* or *character vowels* mark a variety of distinctions in different morphosyntactic contexts. The precise function of pre-radical vowels can only be identified in conjunction with the TAM, agreement, and voice properties of the verb form.

Georgian letter	*Phonetic equivalent*	*Harris 1981*	*Present transliteration*
ა	a	a	a
ბ	b	b	b
გ	g	g	g
დ	d	d	d
ე	e, ɛ	e	e
ვ	v, w, f	v	v
ზ	z	z	z
თ	t^h	t	t
ი	i, ɨ	i	i
კ	k’	ḳ	k’
ლ	l	l	l
მ	m	m	m
ნ	n	n	n
ო	o	o	o
პ	p’	ṗ	p
ჟ	ž	ž	zh
რ	r	r	r
ს	s	s	s
ტ	t’	ṭ	t’
უ	u	u	u
ფ	p^h	p	p
ქ	k^h	k	k
ღ	ɣ	ɣ	ɣ
ყ	q̣	q	q’
შ	š	š	sh
ჩ	č	č	ch
ც	c (ts)	c	c
ძ	dz	ʒ	dz
წ	ç (t̩s)	ç	c’
ჭ	č̣	č̣	ch’
ხ	x	x	x
ჯ	ǰ	j	j
ჰ	h	h	h

Table 1.2: Transliteration

1.4.1 Morphotactic Preliminaries

A simplified Georgian verb template is in (1); cf. (Hewitt 1995b).

(1) (Preverb$_1$)-(Pron1$_2$)-(**PRV**$_3$)-root$_4$-(Thematic Suffix$_5$)-(TAM$_6$)-(Pron2$_7$)

Pre-radical vowels appear between the prefixal pronominal markers and the root. They do not co-occur with each other and appear in the same slot before the verbal stem. Possible pre-radical vowels are *i-*, *u-*, *e-*, *a-*. The central function of pre-radical vowels is to mark various properties known as *version*, which is inextricably connected with the TAM in Georgian and the division of verbs into roughly valence-based conjugation classes. Both these systems are described in Chapter 3.

Pre-radical vowels comprise a single distribution / form class but have a variety of synchronically and diachronically distinct functions. These can be viewed in two main groups. The first group includes several kinds of semantically motivated marking of participant affectedness or salience: 'objective,' 'subjective,' and 'locative' version (with active verbs); the absence vs. presence of an affected participant (with passive and unaccusative verbs); and patient agreement (with some active and unergative verbs).

The second group includes more grammatical functions: conjugation type (with unaccusative verbs); perfect and pluperfect paradigm marker (in inversion constructions); future paradigm marker (with unergative and some unaccusative verbs); and several purely morphological functions like the marking of a lexical class.

Some functions of version vowels are historically related. In other cases, different version vowels may have become morphologized in the same slot, but were never part of a single functional category. That is, there may have been two separate grammaticalizations of the same phenomenon, or a secondary conflation of two originally distinct forms.

1.4.2 Prototypical Uses of Version: Active Constructions

The most prototypical uses of version vowels occur in active constructions (transitive and unergative intransitive verb classes). Here, version vowels indicate that some discourse participant is particularly affected by the action (so-called "primary affectedness" or salience, in addition to the action's effect on the patient). The affected participant is a beneficiary or maleficiary; it can be the grammatical subject ("subjective version") or the indirect object ("objective version"). The expression of version is optional and discourse

(or pragmatically) motivated.

Objective version

With objective version, the participant affected by an action is encoded as an indirect object and the verb has an added version marker. The version vowel depends on the person of the indirect object: for 1st and 2nd person indirect objects, the vowel is *i-* (2b). For 3rd person indirect object, the vowel is *u-* (3). In the first case, the pre-stem agreement slot is occupied by object markers referring to the indirect object. In the second case, the pre-stem agreement slot is empty or contains a subject marker.

(2) a. *meri-m* *da-xat'a* ***chem-tvis*** *surat-i*
Mary-NAR PV-paint.AOR **1sg-for** picture-NOM
'Mary painted a picture for me.[2]'

b. *meri-m* *da-**m-i**-xat'a* ***(me)*** *surat-i*
Mary-NAR PV-1SGOBJ-**prv**-paint **(1sg)** picture-NOM
'Mary painted a picture for me.'

(3) *meri-m* *da-**u**-xat'a* *p'avle-s* *surat-i*
Mary-ERG PV-**ver**-paint.AOR Paul-DAT book-NOM
'Mary painted a picture **for Paul**.'

The syntactic function of objective version is in some instances very similar to that of an applicative (Peterson 1999). However, once the full extent of construction in which version vowels participate has been examined in Chapters 4 and 5, it should become clear that there is only partial overlap between version and the typical applicative function.

Subjective version

Subjective version is indicated by the version vowel *i-* and means that the participant affected by the action is the subject. The meaning is self-benefactive, i.e. the beneficiary is the same as the subject. Boeder (1968) and others have suggested that subjective version also adds an empty indirect object which is co-referenced with the subject; this point will be discussed further in Chapter 4.

[2]Unless otherwise indicated, all data are from native-speaker elicitation by the author.

(4) a. *(me) saxl-s* ***v-i**-shen-eb* **mas*
(I) house.DAT **1Subj-ver**-build-TS *he.DAT
'I build a house **for myself**.'

b. *meri saxl-s* ***i**-shen-eb-s* **mas*
Mary.NOM house.DAT **ver**-build-TS-3SGSUBJ *he.DAT
'Mary builds a house **for herself**.'

Objective and subjective version can be indicated by the same vowel, but their functions are clearly distinct. However, the only way to distinguish them formally is by looking at the agreement system and at the arguments cross-referenced on the verb. In other words, it is the particular formal constructional combinations that yield the appropriate semantic interpretations.

Locative version

Active constructions in Georgian permit one more type of version, the so-called 'locative' ('superessive') version expressed by the vowel *a-*. This again indicates affectedness of an indirect object, but further specifies that the action is done in some spatial relation to the indirect object (usually 'onto').

(5) *me* ***v-a**-t'ex-av* *j'ox-s* *(mis)* ***tav-s***
I **1Subj-ver**-break-TS stick-DAT (3SG.GEN) **head-dat**
'I break a stick **over** (his) head.'

1.4.3 Further Grammaticalized Uses of Version Vowels in Georgian

Version vowels in Georgian are used in other constructions to serve functions that are quite different from their prototypical uses.

Unaccusative / Passive

In the passive constructions, version vowels in the Georgian verbal template express the *presence* of an affected participant (*-e-*, as in (6b)) vs. its *absence* (*-i-*, as in (6a)). In these constructions, a version marker is required.

(6)

a. ***i***-*c'ereba*
ver-write.INTR.PRES
'It is being written.'

b. ***e***-*c'ereba* *(mas)*
ver-write.INTR.PRES (he.DAT)
'It is being written **for/in front of/on him**.'

As with the active constructions, version markers express the general notion of participant affectedness. However, their specific function is determined by the larger construction. This difference cannot be ascribed to independent meanings of the version markers, but rather must be described in terms of the differences between active and passive formations.

Inverse Constructions

In so-called 'inversion' paradigms (perfect / evidential; cf. Harris 1981, 117), the version vowels are required and serve to distinguish TAM paradigm and person of the subject. In the Perfect, the vowel *i-* is used with 1st and 2nd person subjects (7a, 7b), and *u-* is used with 3rd person subjects (7c).

(7)

a. *(turme)* *saxl-i* *a-**m**-**i**-sheneb-ia.*
(apparently) house.NOM PV-**1sgObj**-**ver**-build-PERF
'I have (apparently) built a house.'

b. *(turme)* *saxl-i* *a-**g**-**i**-sheneb-ia.*
(apparently) house.NOM PV-**2Obj**-**ver**-build-PERF
'You have (apparently) built a house.'

c. *mas* *(turme)* *saxl-i* *a-**u**-sheneb-ia.*
3SG.DAT (apparently) house.NOM PV-**ver**-build-PERF
'(S)he has (apparently) built a house.'

In the Pluperfect, *-e-* is used in all forms (8).

(8)

a. *saxl-i* *unda* *a-**m**-**e**-sheneb-i(n)a.*
house.NOM must PV-**1sgObj**-**ver**-build-PLUPERF
'I was supposed to have built a house.'

b. *saxl-i* *unda* *a-**g**-**e**-sheneb-i(n)a.*
house.NOM must PV-**2Obj**-**ver**-build-PLUPERF
'You were supposed to have built a house.'

c. *mas saxl-i unda a-**e**-sheneb-i(n)a.*
3SG.DAT house.NOM must PV-**ver**-build-PLUPERF
'(S)he was supposed to have built a house.'

Again, the function of the pre-radical vowels may be historically related to active constructions, but synchronically it is quite different and can only be described with reference to TAM distinctions.

Other uses of version vowels

The vowel *a-* is used in the formation of causatives (9).

(9)
a. *v-q'ep*
1SUBJ-bark
'I bark.'

b. *v-**a**-q'ep-**eb***
1SUBJ-**ver**-bark-**ts**
'I make him bark.'

The same pre-radical vowel *a-* and thematic suffix *-eb* also appear in verbs formed from nouns and adjectives. In this case, *a-* is required even when there is no indirect object. Historically, the use of *a-* may signal the change from static to dynamic (from noun to a verb), but synchronically it is simply part of the verb form. This class now contains many verbs that are not denominal. In such verbs, *a-* can often be replaced by other vowels to express objective or subjective version (10c, 10d).

(10) a. *saxl-s **a**-shen-**eb***
house-DAT **ver**-build-**ts**
'You build a house (beneficiary not specified).'

b. *saxl-s **a**-shen-eb mis-tvis / *mas*
house-DAT **ver**-build-TS he.GEN-for / *he.DAT
'You build a house for him.'

c. *saxl-s **u**-shen-eb *mis-tvis / mas*
house-DAT **ver**-build-TS *he.GEN-for / he.DAT
'You build a house for him.'

d. *saxl-s m-**i**-shen-eb *chem-tvis / me*
house-DAT 1SgObj-**ver**-build-TS *1SG.GEN-for / 1SG
'You build a house for me.'

This latter use provides a paradigmatic contrast with objective-version uses of *i-*, and has been analyzed as indicating 'neutral version' (cf. Boeder 1968). However, as can

be seen, the use of *a-* in these contexts is not related to 'real' version expressing primary affectedness, and calling it 'neutral' is misleading, particularly since it is only required in one lexical class of verbs (cf. (11) for a contrasting verb class).

(11) *surat-s* *xat'-av* / **a-xat'-av*
picture-DAT draw-TS
'You paint a picture.'

To summarize, version vowels belong to a distribution class which is not a coherent functional class. Instead, the function of version vowels in Georgian is mainly determined by larger morphosyntactic constructions. In the case of active verbs, the version vowels express discourse- and semantically- based primary affectedness of event participants and can indicate a large number of distinctions. In other constructions (e.g. passive or inverse), the morphological tools by which version can normally be expressed are used to mark other contrasts. The function of version vowels in these contexts is often related to, but more restricted than, the prototypical uses, and may reflect (perhaps most likely) separate or secondary developments. Overall, version in Georgian is sensitive to both semantic / discourse factors and the morphosyntactic and lexical properties of the verbs.

Chapters 4 and 5 of this study examines the types of contexts in which version is used, its function and distribution.

1.5 Parallels to Version in Other Languages

Although the term 'version' originates in the Kartvelianist tradition, the functional category(ies) that it represents are by no means unique to Georgian and related languages.

Anderson and Gurevich (2005) describe several cases of version from elsewhere in Eurasia, including Turkic, Munda, and Burushaski, and suggest some possible historical trajectories for the development of version markers. While the historical development of version is not the focus of this dissertation, it is useful to look at the types of contrasts marked by version in these other languages, as well as the ways in which version interacts with other aspects of the particular language.

The category of affectedness marking appears to be fairly similar across the languages examined in this section. At the same time, the expression of version is in a sense 'parasitic' on other morphosyntactic constructions of a language, and thus quite different

from one language family to another. In Georgian, as argued in chapters 4-5, version is tightly coupled with the TAM system, valence and voice alternation constructions, and verbal agreement; in Turkic, it is expressed via auxiliary verb constructions; in Burushaski, via a verbal prefix; in Gorum, via a verbal suffix and a suprasegmental creaky-voice feature. A broader pattern seems to suggest that historically, version markers may derive from deictic verbs meaning 'come' and 'go,' but much further research is needed to confirm this hypothesis on a larger scale, and no historical evidence of this sort is available for Georgian.

Most of the data in this section come from (Anderson and Gurevich 2005).

1.5.1 Turkic

Outside of the Caucasus, Turkic has perhaps the best-documented set of version contrasts, marking both *subject(ive)* and *object(ive)* versions (Anderson 2001, 2004). In this language family (see Table 1.3), there are a number of different formal constructions used to mark subject version, all involving a functional specialization of the verb 'take' (**al*). Such formations, depending on the theoretical persuasion of the investigator, may be labeled 'auxiliary,' 'serial,' or 'light verb' constructions.

Language	**LV form**	**AV stem**	**Citation**	**Gloss**
Tuvan[i]	-Ip	al	*biž-**ip** **al**-di-m*	'I wrote (down) for myself'
Tofa	-GAʃ	al	*tùt-kaʃ al-ɣan*	'caught for himself'
Yakut (Sakha)	-An	ɨl	*taay-an ɨl-la*	'he guessed for himself'
Chuvash	-sA	il	*kălarsa il*	'steal for self'
Xakas	-(p)	-ɨl	*tab-ɨl-za-m*	'if I find myself'
Uyghur	-iw-	-al-	*yez-iw-al-di-m*	'wrote down for self'
Xalaj	-Ø-	-al-	*tut:-ā́l-du-m*	'I seized it (for myself)'
Orkhon Turkic	α	al-α	*ölür-tü-müz al-tɨ-miz*	'we killed for ourselves'

(Sources: Anderson and Harrison 1999; Rassadin 1994: 198; Korkina 1982: 289; Skvorcov 1999: 111; Field Notes [GDSA]; Hahn 1991: 612; Doerfer 1988: 169; von Gabain 1974: 279 l.3)

i. so too Tofa, Tuvan, Xakas, etc. LV = lexical verb; AV = auxiliary verb

Table 1.3: Subject version auxiliary verb constructions in selected Turkic languages

The inflectional patterns and the historical morphosyntactic relation between the two verbs in the version constructions vary across different Turkic languages. Lexical verbs in these version 'auxiliary' constructions include various types of converb forms (*-Ip*, *-An*, *-sA*, depending on the language); same subject form (one Tofa variant); or doubled inflection (Orkhon Turkic).

The version constructions in the Turkic languages are usually synchronically bipartite, consisting of a lexical verb and a functional verb that marks version, but may show complete or lexically restricted tendencies to univerbation (Uyghur and Xakas, respectively).

A full example of subject version may be seen in (12) from Xakas:

(12) XAKAS

*pu kniga-nɨ tab-**ɨl**-za-m min xayda örɪn-e-m*
this book-ACC find-**subj.vers**-CON-1 I oh.boy be.happy-FUT-1
'If I find myself this book, boy will I be happy.'
(field notes [GDSA], cited in (Anderson and Gurevich 2005))

Some more detailed examples of Turkic subject version are given below[3].

(13) TUVAN

a. *biž-**ip** **al**-di-m*
write-**cv** **subj.vers**-PAST.II-1
'I wrote (it) (down) for myself'
(Anderson & Harrison 1999: 68)

b. *süt-ten iž-**ip** **al***
milk-ABL drink-**cv** **subj.vers**
'drink some milk'
(Anderson & Harrison 1999: 68)

(14) YAKUT

*kurdarɨ taay-**an** **ɨl**-la*
immediately guess-CV SUBJ.VERS-PST
'he guessed himself immediately'
(Korkina et al. 1982: 289)

[3]I wish to thank Greg Anderson for pointing me to these examples

(15) CHUVASH

a. *astu**sa*** ***il***
remember-**cv subj.vers**
'remember'
(Skvorcov 1985: 111)

b. *kălar**sa*** ***il***
steal-**cv subj.vers**
'steal for self'
(Skvorcov 1985: 111)

(16) TOFA

dilɣi oluk *bariɨp* *brœœ* *yʃpyl* *tùt-**kaʃ*** ***al**-ɣan.*
fox right.away go-CV one hazel.grouse catch-**ss subj.vers**-pst
'Right away the fox caught a hazel grouse'
(Rassadin 1994: 198)

(17) UYGHUR

a. *adris-i-ni* *yez-**iw**-**al**-di-m*
address-3-ACC write-**cv-subj.vers**-PST-1
'I wrote me down her address'
(Hahn 1991: 612)

b. *qol-um-ni* *kes-**iw**-**al**-di-m*
hand-1-ACC cut-**cv-subj.vers**-PST-1
'I got me cut on my hand'
(Hahn 1991: 612)

(18) ORKHON TURKIC [Kül Tegin]

ölür-tü-müz ***al**-ti-miz*
kill-PST-1PL **subj.vers**-PST-1PL
'we killed us...'
(von Gabain 1974: 279 l.3)

(19) XALAǰ

tut:-ā́l-du-m
hold **subj.vers**-PST-1

'I seized (it)'
(Doerfer 1988: 169)

Object version, though less frequent, is also attested in Turkic, using a similar complex predicate construction with the verb 'give'. An example of this formation may be seen in Tofa (20).

(20) TOFA

onu *sooda-p* *beer* *be*
s/he.ACC say-GER **obj.vers**.P/F Q
'Should I say it (again for you)?'
(ASLEP Field Notes [PVB], cited in (Anderson and Gurevich 2005))

Despite the range of formal ways of marking version, the functional category itself seems to be fairly robust across Turkic. Subject version constructions indicate a sense of self-directedness or self-affectedness, similar to the function of subjective version in Georgian.

1.5.2 Burushaski (isolate)

In the Himalayan language Burushaski, a verbal prefix participates in a number of version-like and valence alternation constructions, suggesting a similar kind of symbiosis to that found in Georgian.

The so-called D-prefix (Anderson in press) is a lexicalized part of certain stems. It occupies position –3 in the Burushaski verb template, often appearing with a following copy vowel (from the following syllable) or with a following *-i-*. Anderson (to appear) suggests that this prefix represents an advanced stage in a now lost system of grammaticalized version. Its semantics range from a clear cislocative meaning, to subject version or actor/subject focus (Bashir 1985), to vaguely (de-)transitivizing functions and various idiosyncratic, sometimes opaque semantic nuances.

Verbs with the D-prefix can be both transitive and intransitive. Some examples with this prefix are in (21)

(21)

di-yaraẏ *d-yuraẏ*
'be hot' 'make hot, heat'
(Berger 1998: 107)

d-:s-karaẏ *-r > d-r* *su ∼ dusu-*
'send' 'send here' 'bring'

di-s-íl (∼ di-íl) *du-s-úlǰa (∼ d-úlǰa)*
'become wet' 'satisfy, sate'
(Berger 1998: 109)

1.5.3 Gorum(Munda)

Another language discussed in (Anderson and Gurevich 2005) is Gorum (Parengi), a South Munda language spoken in southern Orissa, India. In Gorum, there are two elements with somewhat overlapping functions, both of which are described as marking 'orientation' or 'affectedness'. Formally, the two elements are quite distinct: one is the suffix *-ay* and the other is a suprasegmental creaky voice feature.

Creaky Voice Feature

One function of the creaky voice feature is to encode an affected subject, i.e. subject[ive] version (22).

(22) $\Longrightarrow$ Subject affecting

GORUM

a. *b. g̰osaN ga?a-ru-ni*
B rice eat-PAST-DESIA.ORIYA.PROG
'B is eating the rice' [4]

b. *g̰osaN ga?a-rṵ--ni*
rice eat-PAST-**ver**-DESIA.ORIYA.PROG
'B is eating himself the rice'

Besides so-called 'subject version' functions, such formations in Gorum may accompany actions primarily affecting a (deictic/speech-act participant) non-subject, and a range of functions generally subsumed under the grammatical heading of 'voice', e.g. 'passive potential', 'antipassive' and 'semi-reflexive'. Thus, unlike Turkic and Georgian, a single version construction in Gorum may mark all types of participant affectedness (sometimes grammaticalized in combination with particular lexemes as part of a construction encoding various modal, aspectual, and/or voice categories, as in Georgian).

(23) $\Longrightarrow$ Object affecting

GORUM

a. *bo?y amon e-niN bam-(m)-ḭN duk-ḭN*
one arrow OBJ-1 hit-1OBJ.**ver** AUX-1OBJ.**ver**
'(an arrow) has hit me'
(Aze 1973: 298)

[4] desia.oria.prog = progressive morpheme borrowed from Desia Oriya

b. *tileʔy-di* *neʔ* *zel-om* *tay-om-t-o͡m* *dom*
old.man-FOC it tell-2 AUX-2-NPAST-2.**ver** MOD
'it's the old man, let him tell you'
(Aze 1973: 278)

(24) ⟹ Passive Potential

GORUM

a. *aɽi* *ʈableʈ* *luʔm-tu͡*
that tablet swallow-NPAST.**ver**
'that tablet will/can be swallowed'
(Aze 1973: 286)

b. *aɽi* *ʈableʈ* *lor-luʔm-nu͡*
that tablet NEG-swallow-INTR.**ver**
'...will/can not...'

(25) ⟹ Antipassive

GORUM
noʔn *turyaʔ-n-u͡* *luʔru*
he kick-INTR-INF.**ver** AUX
'he is kicking (in general)'
(Aze 1973: 284)

(26) ⟹ 'semi-reflexive' or 'auto-affective'

GORUM

a. *miN* *siʔ-doy* *ne-poʔ-tu*
I arm-3 1-stab-NPAST
'I will stab his arm'
(Aze 1973: 281)

b. *miN* *siʔ-niN* *ne-poʔ-tu͡*
I arm-1 1-stab-NPAST.**ver**
'I will stab my arm'
(Aze 1973: 281)

Suffix *-ay*

The cislocative/speaker-orientation suffix *-ay* (from 'come') in Gorum marks motion towards the speaker, orientation toward the speaker, and by extension, motion / orientation toward the 'locus of discourse focus' or deictic center. In some instances, it seems also to have taken on a meaning of a first person singular referent; the same is true of a cognate element in certain conjugations in its sister language, Sora.

(27) GORUM

a. *le-ɖon-tu*
1PL-take-NPAST
'we will take'
(Aze 1973: 275)

b. *le-ɖon-t-**ay***
1PL-take-NPAST-**cloc**
'we will bring'
(Aze 1973: 275)

c. *ara-di* *liga?n-t-***a͡y** *ua?-t-ay*
tree-FOC break-NPST-**cloc.ver** AUX-NPST-CLOC
'the tree will break (towards me)'
(Aze 1973: 279)

In some uses, the 'orientation' seems to act more like the discourse-based notion of salience or primary affectedness, and thus akin to the function of the suprasegmental creaky voice feature.

(28) GORUM

a. *guro?y* *do?-r-iN*
shy.AFF feel-PAST-1
'I felt shy'
(Aze 1973: 275)

b. *guro?y* *do?-r-iN-***ay**
shy.AFF feel-PAST-1-**cloc**
'I felt (myself) shy'
(Aze 1973: 275)

c. *saybu-di* *zel-iN*
master-FOC tell-1
'the master told me'
(Aze 1973: 275)

d. *saybu-di* *zel-iN-***ay**
master-FOC tell-1-**cloc**
'the master told me (and I was affected by that)'
(Aze 1973: 275)

Based on parallels with historical developments in Burushaski, Anderson and Gurevich (2005) suggest that the suffix *-ay* may be synchronically overtaking functions of the older creaky-voice feature. Importantly, both markers are involved in marking version as well as a number of voice- and valency-related alternations.

1.5.4 Ethical Dative

There are also functional parallels between version and the so-called *ethical dative*, attested in Latin and fairly common in Slavic, particularly South and West Slavic. The ethical dative "is a pragmatic means by which the speaker includes the event in the personal sphere of the hearer, so that he/she is included in the narration of the event." (Šarič 2003). Some examples are given in (29-31).

(29) Ethical dative in Serbo-Croatian

a. *Oni su ti* *pravilno odgovorili Mileni.*
they are you.DAT correctly answered Milena.DAT (you=ethical dative)
They did the right thing in answering Milena.
They gave Milena a correct answer.

b. *Juče sam ti joj pomogla.*
yesterday am you.DAT her.DAT help (you=ethical dative)
Yesterday, I helped her.(Bošcovič 2004)

(30) Ethical dative in Czech

a. *Oni* ***ti*** *se mi ani neomluvil.*
he you.DAT refl.ACC me.DAT neg-even apologized
'(I'm telling you), he didn't even apologize to me.' (Fried 1994, 173, cited in Franks and King 2000, 110)

b. *Tak som* ***ti*** *mu pomohol.*
so aux.1SG you.DAT him.DAT helped
'So, listen / would you believe it, I helped him.' (Franks and King 2000:131)

(31) Ethical dative in Russian

a. *Ja tebe ne mal'chik, ja nemolodoj chelovek.*
I you.DAT not boy I not-young person (you = ethical dative)
'I'm not a boy, I'm an elderly person.' (from the Russian National Corpus)

b. *Zdes' vam ne ravnina, zdes' klimat inoj.*
here you.DAT not valley, here climate different (you=ethical dative)
'Here is no valley, the climate here is different.' (from a song by V. Vyssotsky)

The Slavic ethical datives are very frequently second-person, as a means of involving the addressee in the events being described. The formal similarity between ethical dative and version is in the inclusion of a dative-marked pronoun or clitic which would not otherwise be expected. However, in contrast to the situation in Georgian, Franks and King (2000) do not consider ethical datives to be syntactic arguments: there is nothing else in the sentence that those clitics may refer to, and they follow different ordering patterns than argument datives. In addition, ethical datives mark only a very general category of affectedness, without the finer distinctions possible in Georgian and Turkic.

Finally, ethical pronouns have been noticed in old Indo-European languages, as well. Yakubovich (2006) describes the Hittite reflexive particle *=z(a)* as an 'ethical reflexive', akin to subject-oriented version found in Georgian. This clitic seems to be less sensitive to conversation participants (as the Slavic ethical datives are) and more sensitive to grammatical relations, placing it functionally somewhere in between the Slavic cases and the Georgian one.

The parallels to version from other languages demonstrate that marking participant affectedness is not unique to Georgian, and that such marking typically interacts closely with other features of a language's morphosyntax. The impossibility of separating 'core version' marking from other features supports the need to account for such systems in a unified constructional framework. The next chapter introduces such a framework, called here Constructional Morphology. The following chapters (4-6) demonstrate how this framework can be used to describe a complex formal and functional category (such as version) along with the larger morphosyntactic system into which it is integrated. Extensions of the framework to cross-linguistic phenomena that are related to version structurally rather than functionally will be examined in Chapter 7.

Chapter 2

Theoretical Framework

2.1 Introduction

A central aim of most contemporary linguistic theories is to identify the grammatically significant patterns or units in a language, produce a representation for them, and provide a mechanism by which existing and novel linguistic items are defined with reference to such patterns. Theories differ in what they take as their operative units and how they describe the relationships between them.

One major division concerns the way linguistic structures are constructed. Many models (such as morpheme-based theories of morphology) assume that meaning is carried by small linguistic units which are then composed into larger structures; the meaning of the larger structures is a function of the meanings of the smaller parts, or is constructed "bottom-up." Other models (such as construction-based approaches to syntax and word-based approaches to morphology) propose that the smallest units of form need not also be the smallest units of meaning; rather, larger structures may determine "top-down" both their overall meaning and the selection of smaller units within them.

The goal of this chapter is to synthesize recent results from cognitive linguistics and morphological theory into a coherent theoretical framework capable of representing the patterns inherent in the Georgian verbal system. The proposed framework draws on insights from Construction Grammar (CxG) and Word-and-Paradigm (WP) approaches to morphology, described below. In a sense, this theoretical framework revives many of the pre-generative intuitions about the role of phrases and whole words in the description and use of grammar. The following chapters on Georgian aim to demonstrate that these

intuitions provide a natural, as well as psycholinguistically plausible, way of capturing recurrent patterns.

The traditional intuition behind the use of constructions and words is that they provide the most useful basis for syntactic and morphological generalizations. Such approaches are particularly well-suited for representing established patterns. By contrast, bottom-up approaches such as morpheme-based approaches to morphology and many generative approaches to syntax focus on the way in which novel items can be used and novel linguistic structures can be formed. The rest of this chapter provides arguments in favor of a top-down approach to grammatical generalizations, and suggests that analogy, rather than the application of completely abstract general rules, plays a crucial role in how new linguistic elements are formed.

A less central purpose is to outline a formalism for representing the relevant morphosyntactic patterns. For the sake of convenience and readability, the formalism developed in this chapter will use argument-value matrices (AVMs) to represent sets of feature-value constraints on constructions. AVMs are widely used in lexicalist theories of grammar, such as Head-Driven Phrase Structure Grammar (HPSG; Sag et al. 2003) and Lexical-Functional Grammar (LFG; Kaplan and Bresnan 1982, Dalrymple 2001). AVMs are also used in some recent work on Construction Grammar (Kay 2002, Fillmore et al. forthcoming, (Michaelis and Ruppenhofer 2001)). The feature structures are most useful for representing argument structure constraints, valence alternations, and the relation between argument structure and morphological structure. Generalizations about paradigm structure will be approached from a more traditional perspective and represented as links between related forms.

2.2 Constructions and Construction Grammar

2.2.1 Background

Until comparatively recently, nearly all approaches to syntax have been explicitly or implicitly 'construction-based.' From classical grammars to 'standard-theory' transformational grammars, grammatical organization at the level of phrases and clauses is analyzed in terms of larger patterns with a characteristic form, meaning and usage. The principal challenge to this traditional conception comes from the 'atomistic' approach of (Harris 1951) and later generative models that set out to decompose constructions into conspiracies of

conditions (Chomsky 1977) or principles (Chomsky 1991, 1995).

Harris' program was motivated by the desire to account for complex grammatical structure through the interaction of simple operations (Matthews 1999), an idea that survives as early transformations in (Chomsky 1957). Later generative models of syntax were driven by the ideal of "uniformity", meaning that related forms (like the active and passive) must start out the same at some level and get their surface form through regular derivations (Culicover and Jackendoff 2005). The locus of uniformity was thought to be syntactic, with variations in semantics and phonological form deriving from it. Once the derivations were completed, however, these models expected a perfect one-to-one match between form and meaning. Along the way, the intuition that phrases and words provide the most stable mapping between form and meaning was lost or deemed unimportant.

Fillmore's 'Case Grammar' (Fillmore 1968) was one of the earliest attempts within the generative tradition to acknowledge the importance of semantics and semantic roles, and to start moving away from the syntax-centered view of language. Case Grammar proposed that the deep structure of a sentence contains a predicate and a set of case meanings, which are then mapped onto grammatical roles in the surface structure via various mapping rules. The meaning of the sentence is determined by the deep case roles rather than by the syntax. By starting with the meaning, Case Grammar re-introduced a kind of meaning-to-realization, top-down approach, as opposed to the bottom-up construction of meaning from the meanings of the parts assembled in syntactic structures.

Construction Grammar is a family of related approaches rather than a unified framework. In addition to the importance of semantics, the model of 'Construction Grammar' narrowly construed (hereafter 'CxG') was motivated by the desire to account for all aspects of grammar, in opposition to transformational grammar's emphasis on "core grammar." Thus CxG aims to represent all types of linguistic expressions, from extremely productive and abstract syntactic rules to lexicalized idioms, as "constructions all the way down" (Goldberg 2003).

The traditional premise of CxG is that the basic unit of linguistic analysis should be a *construction*, rather than a single morpheme or a word. A construction is a pairing between form and meaning (similar to the Saussurean sign) that can exist at a sub-word, single-word, or multi-word level. CxG's main departure from the generative models of grammar is a looser notion of compositionality. The meaning of a given construction does not have to be strictly composable from the meaning of its parts; rather, the combination of

the parts can add to the individual meanings of the parts or can override them altogether, and this will be specified as a new construction at the relevant level. The flow of information in such a model is from the top down: from the larger construction to the selection of specific sub-parts.

2.2.2 Varieties of 'Construction Grammar'

Construction Grammar has often been used to represent the meanings of semi-productive idioms such as *let alone ...* (Fillmore et al. 1988) or *What's X doing Y* (*WXDY*) (Kay and Fillmore 1999). In these expressions, the syntactic elements of the construction are not sufficient to account for the entirety of its meaning.

For example, the sentences in (32) exemplify the so-called *WXDY* construction, which carries the implication of surprise or indignation that cannot be derived from any of its individual elements.

(32) (Kay and Fillmore 1999:3)

a. What is this scratch doing on the table?

b. What do you think your name is doing in my book?

c. I wonder what the salesman will say this house is doing without a kitchen.

d. I don't suppose the police had anything to say about what their so-called detective thought the footprints were doing under the bedroom window.

e. What is it doing raining?

f. 'What's a nice girl like you doing in a place like this?'

g. What am I doing reading this paper?

In addition to the pragmatic markedness, this construction is characterized by containing the verb DO in the present participle form; this verb may not be a complement to verbs other than the copula BE; its interpretation is not necessarily that of progressive aspect; and a variety of other syntactic restrictions not normally expected to apply to this verb do not apply. Kay and Fillmore (1999) argue that semi-schematic constructions like *WXDY* are legitimate patterns in language, and can and should be described in the same formal system as more general syntactic and semantic structures.

The same type of pattern extension that permits an analysis of semi-productive constructions applies to more general syntactic combinations as well. Goldberg (1995) has

suggested that even more abstract syntactic patterns should be represented as constructions that carry some meaning independent of the meanings of the parts. Her main objective was to account for argument-structure alternations where verbs not normally observed with a particular argument structure nevertheless make sense when used in a particular context. For example, the verb 'sneeze' would be considered intransitive and not expected to appear with a direct object or a goal prepositional phrase. However, phrases like "She sneezed the foam off the cappuccino" are observed, and are interpretable in the proper context.

Goldberg suggests that argument-structure patterns are themselves constructions and introduce meaning that is not simply a function of the sum of the nouns and verbs in them. In particular, the ditransitive construction has the meaning of 'transfer' that reflects a dynamic scene, independently of or in addition to the elements in it. Moreover, she suggests that such abstract patterns get their meaning in speakers' minds from the individual lexical items that frequently occur in them. The most frequent verb to appear in a ditransitive construction is 'give', as evidenced by corpus studies of both adult speech and mothers' speech directed at children (Cameron-Faulkner et al. 2003), and so the generalized ditransitive constructions carries the meaning of 'transfer.'

The main advantage of CxG is that it is capable of representing form and meaning generalizations independently of each other and at different degrees of abstractness. Anything from a lexical (single-word) construction to an argument-structure construction can be represented in the same format. A given linguistic expression can instantiate (or inherit from) multiple constructions at the same time. The compatibility of the different parent constructions can be guaranteed by the process of *subsumption*, or *constraint satisfaction* as used in Head-Driven Phrase Structure Grammar (cf. Sag et al. 2003) (some versions of CxG use unification instead). Two constructions successfully license a linguistic expression if they do not impose conflicting constraints on it. The use of subsumption to maintain consistency will be presented in more detail below.

In addition, elaborate inheritance hierarchies of constructions can be created to represent sub-generalizations within generalizations. Most current constructional approaches assume default or *normal mode* inheritance (Flickinger et al. 1985), such that more specific constructions inherit from more general constructions and may override or fail to inherit some of their parents' characteristics.

Michaelis and Lambrecht (1996) analyze the English nominal extraposition construction, as exemplified in (33) and suggest that it is set apart from other syntactic pat-

terns by the particular constraints on word order, lack of coreference restrictions between the fronted constituent and the main clause, and the unusual information structure of such sentences.

(33) Nominal Extraposition Construction (from Michaelis and Lambrecht 1996).

a. It's AMAZING the people you SEE here.

b. It's ASTONISHING the age at which they become skilled LIARS.

c. It's STAGGERING the number of books that can pile up.

Despite the uniqueness of this construction, the individual properties that set it apart are attested in other, more general construction types. For example, the fronted word order is also characteristic of Right-Dislocation constructions (e.g. *It's amazing, the things children say*, as well as a number of other constructions. This, Michaelis and Lambrecht claim, is evidence for a more general construction describing *extraposition*. Both the Nominal Extraposition and Right Dislocation constructions inherit from the general extraposition construction. Thus, a variety of formal, semantic, and pragmatic constraints or features can define a construction.

According to Goldberg (2006:215), the major theoretical divisions in construction-based approaches are between the original Berkeley Construction Grammar of Kay and Fillmore, which relies heavily on unification (and hence is termed Unification Construction Grammar), and the more psycholinguistically motivated approaches of Cognitive Construction Grammar (Goldberg 1995, Lakoff 1987), Cognitive Grammar (Langacker 1987a, 1987b), and Radical Construction Grammar (Croft 2001). Table 2.1 exemplifies some of the parameters by which these approaches are the same or different.

Recently, there have been some attempts to develop psychologically motivated, usage-based versions of Construction Grammar that nevertheless have an explicit formalism precise enough for computational modeling. One such approach is Embodied Construction Grammar (ECG; Bergen and Chang 2005), which combines the idea of a construction as a form-meaning pairing with a simulation-based model of language understanding (based on work described in Bailey et al. 1997, Regier 1996, Narayanan 1999). ECG is part of a larger project on language modeling, the Neural Theory of Language (Feldman 2006), which also provides a structured connectionist computational base (Feldman 1988). Some recent psycholinguistic studies demonstrate that mental simulation does, indeed, play an important

	CG, CGxG, RCxG	UCxG
Constructions	Learned pairings of form and function	Learned pairings of form and function
Role of constructions	Central	Central
Non-derivational	Yes	Yes
Inheritance	Default	Default (prev. monotonic)
Usage-based	Yes	Not uniformly
Formalism	Notation developed for ease of exposition only	Heavy focus on unification-based formalism
Role of "motivation"	Central	None
Emphasis on	Psychological plausibility	Formal explicitness; maximal generalization

Table 2.1: Similarities and differences between Cognitive Grammar, Cognitive Construction Grammar, and Radical Construction Grammar on the one hand, and Unification Construction Grammar on the other (from Goldberg 2006:215).

role in language understanding (Bergen et al. 2004, Narayan et al. 2004), and neurally plausible language theories are a promising research avenue. The current ECG formalism focuses mainly on semantic and syntactic patterns, and has not yet been expanded to accommodate morphological representations, however. Its main advantage is in representing complex semantic schemas as well as a computationally tractable representation of major syntactic constructions. Neither of these issues are a central concern of this study, however, and therefore the ECG formalism will not be used here.

Rather than committing to a specific variant or formalism of CxG, I will adopt the basic insights and intuitions and focus on parts of the framework relevant for representing the Georgian data. The guiding principles for such a constructional approach are as follows (mostly following (Kay and Fillmore 1999).

- The basic unit of linguistic description is a construction.
- A construction is a set of constraints (phonological, morphological, syntactic, semantic, and/or pragmatic) on linguistic expressions that license *constructs*.
- A *construct* is a linguistic expression that instantiates one or more constructions.
- Constructions are organized into an inheritance hierarchy.
- This is a default inheritance hierarchy: child constructions are assumed to inherit all

of the properties of their parents unless these properties are explicitly overridden in the child constructions.

- A given lexical expression instantiates one or more constructions. Multiple constructions are reconciled via constraint satisfaction, used here in the same sense as in HPSG grammars. See section 2.4.4 for a discussion of what it means to "instantiate" a construction.
- Constructions are taken to be language-specific, although some more abstract constructions may be common to multiple languages.
- The idea that a language contains multiple, possibly overlapping, patterns at various degrees of generality is taken to be universal and constrained by human cognitive capabilities. The various levels of generalizations can be represented as an inheritance hierarchy, although the hierarchical structure need not (and probably does not) correspond to anything real in the brain.

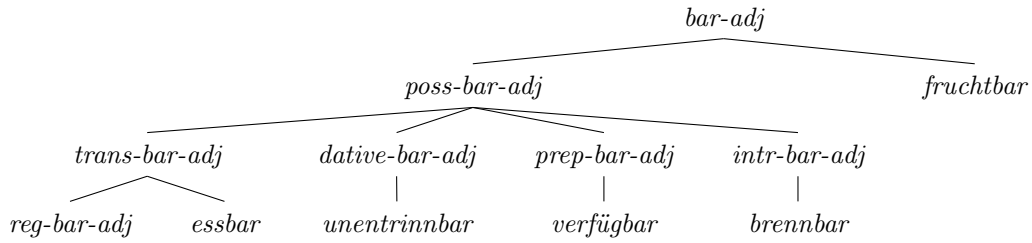
2.2.3 Constructional Approaches to Morphology

Although Construction Grammar started as a theory of syntax and semantics, with particular emphasis on the interaction between the two, there have been some recent attempts to extend it to (derivational) morphology. Riehemann (2001) takes a construction-based approach to derivational morphology in order to account for the formation and productivity of *-bar* and *-able* adjectives in German and English, respectively, and for some non-concatenative derivational patterns in Hebrew. Reihemann's approach relies on a well-elaborated hierarchy of lexical types at various degrees of specificity.

The formation of the German *-bar* adjectives like *essbar* ('edible', from *essen* 'eat') exhibits a number of semi-regular constraints on the semantics, syntax, and phonology of the participating verbs. However, 'there does not seem to be a single generalization that is general enough to encompass all existing and possible *bar*-adjectives and at the same time specific enough to exclude impossible examples' (Riehemann 2001:245). There are general, productive patterns (adjectives derived from transitive verbs); subregular patterns (adjectives derived from verbs with dative or prepositional objects), as well as a number of lexical exceptions to each pattern. Riehemann analyzes such adjectives as forming a type hierarchy in which both the regular and the irregular patterns inherit from the same underspecified type 'transitive *bar*-adjective', as in 34. Inheritance in this case is MONOTONIC,

i.e., no overrides are allowed by the lower nodes. The debate about default vs. monotonic inheritance has not yet been settled, although most current constructional approaches have chosen the former. I will not attempt to resolve it here, as it seems that any hierarchical pattern of subregularities can be expressed in either mode.

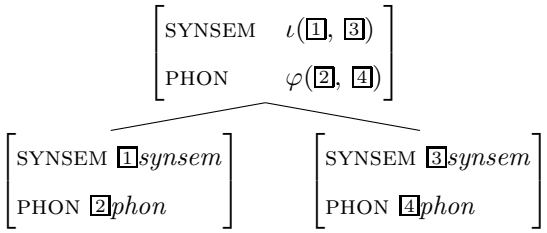
(34) Partial hierarchy for *bar*-adjectives from (Riehemann 2001:264)



Riehemann's approach is limited in that syntactic and morphological patterns are analyzed separately, albeit by a similar mechanism, and there is no treatment of inflectional morphology. The framework proposed in the present study differs from Riehemann's by explicitly dealing with the syntax-morphology interface, and by incorporating inflectional patterns.

Orgun (1996) presents a model of Sign-Based Morphology and Phonology, an approach loosely based on the notion of grammatical constructions; subsequent applications of this framework are in (Orgun 1997, 1999, Orgun and Inkelas 2001), and a similar framework is applied to the phenomenon of morphological reduplication in by Inkelas and Zoll (2005). The basic idea behind Sign-Base morphology is that each node in a syntactic or morphological structure (tree) is a *sign* which imposes a set of form and meaning constraints on its constituents. Constructions such as affixes are defined as local trees, consisting of a mother sign and daughter signs, such that the mother (combination of two smaller elements) can have independent meaning and/or form from the daughters. In such a construction, a stem and affix are daughters, and the resulting word is the mother, as in (35).

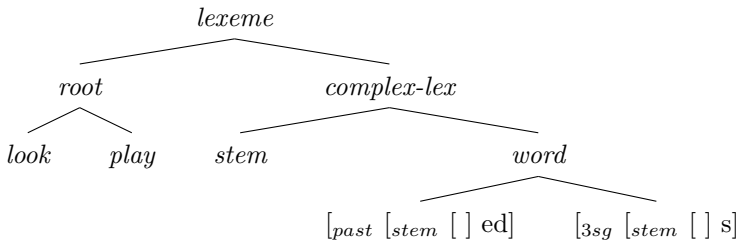
(35) A construction in Sign-Based Morphology (Orgun 1996:73)



Orgun explicitly states that his approach is similar to stem-based approaches to morphology such as (Anderson 1992) and (Aronoff 1994), which grant a special theoretical status to stems. The present approach differs from stem-based morphology in this important respect: as argued in Chapter 3, verb stems alone are not sufficient to predict the formation of Georgian inflectional paradigms, and it is not clear to what extent they are theoretically significant. Another difference is in emphasis: while Orgun focuses on phonology and the phonology-morphology interface, my framework is designed to handle the syntax-morphology interface and inflectional morphology. I do not explicitly deal with phonology, in part because the mapping from morphological to phonological structure in Georgian does not introduce any significant complications.

Koenig (1999) proposes to describe lexical relations in a Type Underspecified Hierarchical Lexicon (TUHL). In such a lexicon, all possible word formations can be described as subtypes of one or more existing categories, so that the production and recognition of novel items is essentially a categorization problem. The lexical categories are constructional and include constructions specifying part of speech types, stems, affixes, and any other morphological elements, as in the hierarchy in 36). The classification of novel words like *looked* consists of constructing a new type, which inherits from both the root *look* and the past-tense construction '[$_{past}$ [$_{stem}$ [ ] ed]'. This happens online for productive verbs (unlike the prevailing assumption in HPSG syntax, where the type hierarchy is 'pre-compiled', or static, prior to any application of grammatical rules).

(36) A simplified hierarchy of morphological types (Koenig 1999:54)



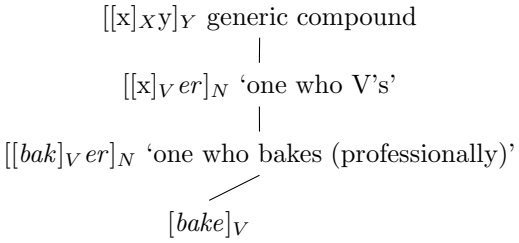
Koenig's approach appears to be somewhere between a realizational and a morpheme-based approach. As in a morphemic approach, roots and affixes have equal ontological status in the hierarchy; however, their full-word combinations, which inherit from both the root and the affix, need not be compositional. Koenig's model allows for these combinations to impose their own restrictions on form and meaning not necessarily derivable from the

parent constructions. He is able to avoid problems with portmanteau morphs and multiple exponence by having separate semantic functions for combining inputs from parent constructions, so that there is no assumption of a one-to-one correspondence of form and meaning. At the same time, this approach is "constructive" in that the form and meaning of the parts is predefined and then put together into whole words. Just as the other approaches described in this section, Koenig's model does not provide an obvious way to accommodate paradigmatic effects or, in general, contrasts between related constructions.

The most recent approach to constructions in morphology is that of Booij (2005, forthcoming). Like Riehemann, he argues for a hierarchical lexicon consisting of schemas of various degrees of abstractness, and provides a model for templatic inheritance.

An example hierarchy for the English or Dutch agentive suffix *-er* is in (37). The *-er* schema inherits from a more general stem+suffix schema. The specific instance of *baker* inherits from the *-er* schema and unifies it with the schema for *bake*.

(37) A hierarchy of compounds (Booij 2005)



There is strong psycholinguistic evidence that this kind of example-based extensio reflects a psychologically real process. Krott et al. (2001, 2002) examine the productive but not fully regular choice of linking morphemes in Dutch compounds and demonstrate that adult speakers' choices are influenced by the distribution of linking morphemes in related words. Such 'morphological' or 'constituent' families of words that share one of the compounding affixes seem to have a strong analogical effect on the production of novel compounds. Subsequent studies demonstrate that the size of such morphological families also influences the way children create novel compounds (Krott et al. forthcoming).

Booij demonstrates why a constructional approach is viable and applies it to Dutch compounding and derivation. The limited scope of applications, however, prevents Booij's framework from being easily applied to the Georgian data considered in this thesis. Again, he does not provide a mechanism for representing the syntax-morphology interactions, and

the format used to represent templates and hierarchies is not yet well-defined, in part because it is not clear that the 'schemas' themselves have any status in the grammar, above and beyond the generalizations that they embody. Booij's approach is in some sense very similar to stem-based approaches: the constructions that serve as input to the derivational schemas seem to associate individual stems with meanings. This may not have been Booij's explicit intent, because in the case of Dutch compounding the stems mostly correspond to full word forms, and so no separate mechanism is needed to abstract away just the stems. However, an artefact of this is that there is no mechanism in Booij's framework for representing meaningless stems.

Thus, the main contribution of Booij's framework is the philosophical foundation for constructional morphology. The approach proposed in this dissertation is not significantly different from that of Booij or Riehemann in its treatment of derivational morphology.

Finally, the present approach focuses on the description of non-compositional patterns and paradigmatic contrasts between related constructions, influenced by the Word-and-Paradigm approaches to morphology.

Most other constructional morphological approaches have focused on individual constructions and not on the relations between them. This focus has worked well for derivational morphology and compounding, areas where paradigmatic contrasts are not often considered necessary for linguistic description. However, as the following discussion will demonstrate, inflectional morphology crucially relies on paradigms and paradigm structure. A construction-based description of inflectional patterns must therefore provide a way to accommodate paradigms.[1]

2.3 Morphology: Words and Paradigms

This section introduces some of the issues often debate in morphological theory and discusses several morphological theories in use in the last century. I begin with a discussion and motivation of *morpheme-based* approaches to morphology that became prevalent in the second half of the century and present some data that challenge the basic principles of those approaches. The following discussion returns to a more traditional model based on *words*

[1]Though see Bochner 1993, Blevins 2001a for arguments that derivational morphology, too, can be considered paradigmatic. The notion of a 'morphological family' likewise corresponds to a traditional derivational paradigm.

and paradigms, whose motivating intuitions account for the challenging data more naturally. The rest of the section looks at previous construction-based approaches to morphology and presents more formally the modern Word-and-Paradigm models.

2.3.1 Background

In the second half of the twentieth century, it became accepted in linguistic theory that the main goal of morphology was to describe the relations between words and their subparts. Although the general approach can be traced to Bloomfield (1933), the development of 'morphemic' models of morphology (based on earlier 'phonemic' models of phonology) was largely due to Harris (1942), Hockett(1947, 1954) and other 'Post-Bloomfieldians'. In such an approach, the smallest unit for descriptive purposes, the morpheme, is also assumed to be the smallest unit of meaning. Whole words are derived from morphemes either through compositional principles, as in Item-and-Arrangement (IA) models (Lieber 1980, Selkirk 1982) or by rule-driven derivation in which morphemes realize specific morphosyntactic properties, as in Item-and-Process (IP) models (Hockett 1954, Steele 1995, Steele 2002).

The key idea behind morphemic approaches is that variation in form is determined by variation in meaning: smallest elements of form correspond to smallest elements of meaning, and so when one changes, so does the other. This assumption was almost a functional necessity for linguists investigating the complex word structure of Native American languages, where morphemes were added up together 'like beads on a string'. The resulting expectation, then, is that elements of form and elements of meaning will stand in a one-to-one correlation with each other.

As a simple illustration, compare the IA and IP treatments of the English plural *-s*. In an IA approach (38a, the plural morpheme, along with its allophonic alternatives, is listed in the lexicon and combined with noun morphemes. In an IP approach (38b, a rule adds the property "plural" to a noun lexeme, along the way adding a phonological affix to the phonology of the noun.

(38) a. IA: NOUN + <PLURAL, z>

b. IP: <[NOUN], ϕ> → <[NOUN, PLURAL], ϕ+z>

However, both types of theories run into trouble in cases where form-meaning associations are not one-to-one. Many familiar European languages provide plenty of examples

where the form-meaning biuniqueness is violated. Hockett (1954) describes several such cases and concludes that, although both types of theories can be changed to handle them, the required changes would violate the principles on which the theories were built, namely that the lexical properties of morphemes are sufficient to account for their combinations and distribution.

The first type of difficulty for morpheme-base approaches is *cumulative expression*, or so-called 'portmanteau' morphs. In the Russian inflectional paradigm in Table 2.2, the inflectional affixes express the meanings of case and number simultaneously, thus violating the one-to-one correspondence of form and meaning that defines morphemes.

	SINGULAR	PLURAL
NOMINATIVE	*ram-a*	*ram-y*
ACCUSATIVE	*ram-u*	*ram-y*
GENITIVE	*ram-y*	*ram-ø*
DATIVE	*ram-e*	*ram-am*
INSTRUMENTAL	*ram-oj*	*ram-ami*
	'frame'	

Table 2.2: Cumulative expression (portmanteau morphs) in Russian

The second type of difficulty involves *zero morphs*, or lack of an overt affix where one would otherwise be expected based on the semantic features. In the Russian paradigm in Table 2.2, the genitive plural form does not have an explicit affix, although all the contrasting forms do have affixes. If the meaning of a word is contributed by its constituent morphemes, then it is not clear how the meaning 'genitive plural' arises in this case. One could speculate that this is the unmarked meaning associated with the root itself, but this is clearly suboptimal since there is no sense in which other person/number combinations could be said to derive their meanings from the genitive plural.

The third type of difficulty is presented by so-called *empty morphs*, or formal elements that do not seem to contribute any explicit meaning. In the Lezgian examples in Table 2.3 (Haspelmath 2002), the non-absolutive noun forms contain an extra suffix in addition to the case ending (*-re*, *-di*, or *-a*). Given the contrast with the absolutive forms, these suffixes are not part of the root; nor are they part of the case meaning, since they stay the same in all cases. Again, this type of phenomenon is problematic for morpheme-based

approaches because the meaning of a morpheme is expected to determine its distribution. Since the Lezgian suffixes do not have identifiable meanings, it is not clear on what basis their distribution can be stated (particularly their absence in the absolutive forms).

ABSOLUTIVE	*sew*	*fil*	*Rahim*
GENITIVE	*sew-re-n*	*fil-di-n*	*Rahim-a-n*
DATIVE	*sew-re-z*	*fil-di-z*	*Rahim-a-z*
SUBESSIVE	*sew-re-k*	*fil-di-k*	*Rahim-a-k*
	'bear'	'elephant'	male name

Table 2.3: Empty morphs in Lezgian (Haspelmath 2002)

Finally, perhaps the most challenging issue for morphemic approaches is so-called *extended* or *multiple exponence*, when a single morphosyntactic property may be marked by more than one morphological element. Consider the Luxembourgish examples in Table 2.4 (Matthews 1991:182).

	SINGULAR	PLURAL
'leg'	*Been*	*Been*
'fish'	*Fësch*	*Fësch*
'brother'	*Brudder*	*Bridder*
'mouse'	*Maus*	*Mais*
'animal'	*Déier*	*Déier*
'table'	*Dësch*	*Dëscher*
'book'	*Buch*	*Bicher*
'house'	*Haus*	*Haiser*

Table 2.4: Extended exponence in Luxembourgish (Matthews 1991: 182)

In the Luxembourgish example, the plural is not marked at all in the first two words; it is marked by a vowel change in the second pair of words; by the suffix *-er* in the third pair of words; and by both the suffix and the vowel change in the last pair. In this case, the form-meaning relationship is many-to-one.

The examples above demonstrate at the simplest level that the assumption of form-meaning biuniqueness is violated in many cases by very common phenomena, and the

discovery methods of finding the smallest units of form may not lead to the smallest units of meaning. The next section introduces several more cases that suggest that paradigmatic comparisons may also be crucial for the interpretation of word forms.

2.3.2 Traditional Word-and-Paradigm Theory

Before the advent of morpheme-based theories, a different view was widely assumed. Both the ancient grammarians and the pre-Bloomfieldian Americanist tradition described relations between whole words within paradigms rather than between morphemes within words (Kuryłowicz 1949b). This approach, known as Word-and-Paradigm (WP), persists even now in pedagogical grammars and has recently been revived in theoretical circles as well (Blevins 2001b, Blevins forthcoming). P. H. Matthews, one of the advocates of reviving the ancient approach, provides the following reasons for it (Matthews 1991:188):

> Firstly, it conforms very closely to the method by which languages of this kind are traditionally taught. Pupils begin by memorising paradigms. [...] They learn that different members of a paradigm are distinguished by their endings. They can then transfer these endings to other lexemes, whose paradigms they have not memorised. [...] Secondly, it is not clear that, when native speakers learn a flectional language, they do not themselves learn words as wholes. [...] [F]inally, both Latin and Ancient Greek had native grammarians; and it is significant that, as native speakers writing for and teaching other native speakers, they too dealt with words as wholes.

Word-and-Paradigm espouses a worldview similar to Construction Grammar, albeit applied to a different domain (morphology rather than syntax and semantics). The main intuition is that the locus of linguistic description should be in the most stable elements (words in morphology and phrases in syntax), which need not be the minimal elements of form, but which appear to carry integral meanings and dictate the selection of their form parts.

By focusing the attention on relations between whole words rather than the meanings of morphemes, WP approaches avoid many of the problems mentioned in (Hockett 1954). In addition, Matthews (1991) mentions another type of phenomena that provides evidence for paradigmatic structure in language. These phenomena involve explicit contrasts between forms in a paradigm. In the Spanish examples in Table 2.5, the indicative and subjunctive are distinguished from each other by the contrast in the suffix vowel: if

the indicative has -*a*, then the subjunctive will have -*e*, and vice versa. Thus, the indicative meaning cannot be attributed to either of the vowels, and neither can the subjunctive meaning.

	INDICATIVE	SUBJUNCTIVE	INDICATIVE	SUBJUNCTIVE
1SG	*compro*	*compre*	*como*	*coma*
2SG	*compr**a**s*	*compr**e**s*	*com**e**s*	*com**a**s*
3SG	*compr**a***	*compr**e***	*com**e***	*com**a***
1PL	*compr**a**mos*	*compr**e**mos*	*com**e**mos*	*com**a**mos*
2PL	*compr**á**is*	*compr**é**is*	*com**é**is*	*com**á**is*
3PL	*compr**a**n*	*compr**e**n*	*com**e**n*	*com**a**n*
	'buy'		'eat'	

Table 2.5: Paradigmatic contrast marking in Spanish (Matthews 1991)

A somewhat more complex problem is presented by so-called *rules of referral* (Zwicky 1985, Stump 1993), where the formation of a form in one paradigm cell depends on the formation of another paradigm cell, regardless of the phonological content of those cells. One such example is the formation of the perfect tenses (Perfect and Pluperfect) in Georgian. As discussed in more detail in the following chapters, Georgian verbs are divided into four conjugation classes based on morphological criteria, although these classes also roughly correspond to syntactic argument structure. The formation of the perfect tenses in Conjugation 1 (transitive) and Conjugation 3 (unergative) is fairly straightforward: the Perfect for many (but not all) verbs is based on the Future stem, and the Pluperfect regularly based on the 3sg Aorist ending, as in Table 2.6 (Aronson 1990, Hewitt 1995b, as discussed in Blevins 2006).

FUTURE	PERFECT	AORIST	PLUPERFECT	GLOSS
damalavs	*daumalavs*	*damala*	*daemala*	'hide'
dac'ers	*dauc'eria*	*dac'era*	*daec'era*	'write'
gaacnob	*gaucvnia*	*gaac'no*	*gaec'no*	'introduce'

Table 2.6: Georgian Perfect Tenses, Conjugations 1 and 3. (Based on Hewitt 1995)

The form interdependence is different, and more striking, in Conjugation 2 (unaccusative) verbs. Here, the Perfect and Pluperfect forms are constructed differently depending on whether the verb has an indirect object (bipersonal, or relative) or not (monopersonal, or absolute). Monopersonal forms are based on the participle forms, while the bipersonal forms are based on the verbal noun (*masdar*), as in Table 2.7.

PARTICIPLE	MONOPERSONAL	MASDAR	BIPERSONAL	GLOSS
dabadebuli	*dabadebuliq'o*	*dabadeba*	*dabadeboda*	'be born'
damaluli	*damaluliq'o*	*damalva*	*damaloda*	'be hidden'

Table 2.7: Georgian Perfect Tenses, Conjugation 2. (Based on Hewitt 1995)

Thus, in some instances, the formation of one form can be predicted by the formation of another form, even if neither form is semantically derived from the other[2].

'Word and Paradigm' as a modern morphological approach was first mentioned by Hockett (1954), although with apologies for not covering it in more depth. Robins (1959) picked up on the idea, suggesting that it be developed in more detail. However, it wasn't until Matthews (1972, 1991) that the approach got off the ground as a theory used for linguistic analysis. It should be emphasized that WP did not develop as a response to inadequacies of morphemic approaches; rather, it was a traditional model with an older underlying intuition, which had been suppressed in the search for minimal 'building blocks'.

The major difference between WP and the other approaches (IA and IP) is the fundamental status which WP assigns to whole words. The essential part-whole relation for WP is that between words in a paradigm, whereas in IA and IP it is between morphemes in a word. Matthews (1972) suggests using WP to deal with flexional languages like Latin and Greek. He argues that WP is better suited to overcome the challenges such languages present for morpheme-based models. These challenges include the types of problems mentioned above as undermining the principles of IA. Instead, Matthews suggests looking towards pedagogical and teaching grammars of flexional languages and designing a method based

[2]The influence of paradigmatically related forms on each other has also been observe in historical morphophonology (Garrett and Blevins forthcoming, Garrett forthcoming). The analogical patterns of change described by these authors are clearly comparable to Word-and-Paradigm morphological models. At the same time, there is no *a priori* reason to equate he importance of analogical paradigmatic comparisons with universal constraints on paradigm uniformity à la McCarthy (2005).

on relations between words within paradigms rather than between morphemes within words.

The contributions of traditional WP models are discussed in detail in (Matthews 1972, 1991) and (Blevins forthcoming). Here, I will briefly summarize some of the major points of the proposals. First, the traditional Word-and-Paradigm model can be thought of as 'abstractive' (Blevins 2006). The rules, or patterns of morphological formation, are abstracted away from full word forms and relations between word forms. An abstractive grammar model is concerned with how patterns can be extracted from data, and the resulting rules have only a secondary status to the patterns that exemplify them. By contrast, Blevins calls the 'building-block' models like Item-and-Arrangement 'constructive', where rules have primary status and the model is concerned with describing how the are put together to produce word forms, which are secondary to the rules. Given the way the human brain receives information (for language, it is as example words and expressions), the abstractive model seems to be quite plausible psychologically, whereas any constructive model must first go through some stage of extracting the rules.

The key component of the traditional WP models is the *exemplary paradigm*. In most flexional languages, inflecting forms are traditionally divided into inflectional classes on the basis of their morphological formation. These classes are best identified using exemplary representatives, which can be analogically extended to the formation of novel items. For instance, Serbo-Croatian nouns are traditionally divided into three declension classes based on the inflectional patterns. These classes are identified as *o*-stem, *a*-stem, or *i*-stem based on the stem structure in Old Church Slavonic, as in Table 2.8[3]. However, in the modern language, many of the stems now have a different final phoneme. Instead, a declension class can be identified by an exemplary paradigm such as the three paradigms in the table. The paradigm of a novel word can be inferred by analogy with an existing paradigm if its declension class is known. The exemplary inflectional classes summarize inflectional variability in a concise manner without having to postulate unmotivated allomorphic variation for the inflectional endings. A relatively small number of inflectional classes, plus a way to identify the class of a new noun, provides for an economical description of an inflectional system.

In addition, paradigms are said to have an *implicational structure*. So, given the exemplary paradigm structure, one (or a few) word forms must be sufficient to assign a

[3]This is a subset of all Serbo-Croatian declensional paradigms; several of the classes can be further subdivided.

	o-stem		*a*-stem		*i*-stem	
CASE	SING	PLUR	SING	PLUR	SING	PLUR
NOM	*prozor*	*prozori*	*žena*	*žene*	*kost*	*kosti*
VOC	*prozore*	*prozori*	*ženo*	*žene*	*kosti*	*kosti*
ACC	*prozor*	*prozore*	*ženu*	*žene*	*kost*	*kosti*
GEN	*prozora*	*prozora*	*ženē*	*žena*	*kosti*	*kosti*, *kostiju*
DAT	*prozoru*	*prozorima*	*ženi*	*ženama*	*kosti*	*kostima*
INST	*prozorom*	*prozorima*	*ženom*	*ženama*	*kosti*	*kostima*
LOC	*prozoru*	*prozorima*	*ženi*	*ženama*	*kosti*	*kostima*
GLOSS	'window'		'woman'		'bone'	

Data from (Browne 1993); accentual marks have been ommitted.

Table 2.8: Serbo-Croatian nominal declensions

given word to a paradigm. This point is crucial for the usefulness of exemplary paradigms: if one had to know the full paradigm for each lexeme in order to assign inflectional class, there would be as many exemplary paradigms as there are nouns, breaking the economy of the description. In the Serbo-Croat paradigms above, the nominative singular form is in most cases sufficient to identify which paradigm a word belongs to. The only exception are consonant-stem nouns, which are ambiguous between the *o*-declension and the *i*-declension. Here, the distinction is cued by gender, as masculine nouns belong to the *o*-declension and feminine nouns belong to the *i*-declension.

The result of describing an inflectional system in terms of exemplary paradigms is that there is no longer an explicit need to break word forms into component parts in search for smallest units of meaning. The 'rules' for constructing words are stored, in a distributed fashion, in the examples, rather than in a separate set of rules (what Blevins 2006 calls the 'separation of data and program'). Following up on this basic intuition, most subsequent WP models have focused on ways of representing the structure of the paradigms and the way word-formation patterns are extended to novel items. In most modern WP approaches, word formation is represented by *realization rules*, or instructions for associating bundles of morphosyntactic properties (paradigm cells) with forms. These approaches are discussed in more detail in section 2.3.4.

2.3.3 Words vs. Morphemes

The first detailed proposal of a word-and-paradigm theory is Matthews (1972, 1991), and more specific implementations have followed (see section 2.3.4). On the other hand, no contemporary IA model is as entirely explicit as the initial models from the 1940s. For now, it will be useful to explicitly compare the basic assumptions of Word-and-Paradigm approaches with those of Item-and-Arrangement approaches.

The most pure modern version of an Item and Arrangement approach is that of Lieber (1980), and I will use it as the basis for comparison with WP. Both types of approaches are interested in part-whole relations between different kinds of elements. For IA, such relations hold between morphemes and words which are composed of morphemes. For WP, the part-whole relationships are between words and paradigms within which words can be contrasted. The definitions of morphemes and words are thus rather different. For IA, the morpheme is a minimal unit of form and meaning, and the word has no special status. For WP, the word is a minimal unit of meaning, even though it may consist of smaller formal units (morphs), and the word has a very privileged status within the system. Matthews (1972:97-98) provides the following arguments for assigning special status to the word, using examples from Latin:

> I. First, the word is the minimal sentence-construction: no group of morphemes smaller than the word could appear or be instanced as an utterance in its own right. [...]
>
> II. Next, the word in Latin would never exhibit a recursive or self-repeating construction. [...] There is nothing in the grammar [...] which would exclude a clause or phrase of indefinite extent and complexity. By contrast a word can never exhibit more than a certain fixed number of elements. [...].
>
> III. Thirdly, the cohesion of morphemes within the word is considerably greater than the cohesion of words within the sentence. In the sentence it is possible, and even normal, for a syntagm to be interrupted by certain extraneous elements [...] A word, on the other hand, can never be interrupted in a similar way [...].
>
> IV. Finally, the patterning of elements within the word is both fixed and non-contrastive in character. At sentence-level it is well known that one could have both *we:nerit tiberius* and *tiberius we:nerit* – with, presumably, a difference in emphasis or conditioning from the discourse context. But again there is no parallel whatever at word-level: one could not have a variation – let a lone a semantic or stylistic contrast – between *tiberius* and *s-tiberiu*, or between *we:nerit* and *it-we:ner*, and so forth! The most that can happen is that a certain morpheme (viz., the Perfective) may show a variation in position from one paradigm to another.

On the other hand, the paradigm is assumed to have no theoretical status in IA theories (Lieber 1980:119). Thus while morphemes in IA are units signaling morphosyntactic properties, the morphs in WP are only exponents of properties and mostly mark contrasts between paradigm cells.

Stump (2001) proposes two dimensions along which morphological theories can be distinguished. The first is lexical vs. inferential theories. *Lexical* theories list all morphemes in the lexicon, roots and affixes alike. *Inferential* theories express relations between roots and full word forms by rules or formulas. The second distinction is between incremental and realizational theories. *Incremental* theories state that a word acquires morphosyntactic properties by adding affixes, i.e. informational content is constructed bottom-up. *Realizational* theories state that a set of morphosyntactic properties licenses a certain combination of morphs, i.e. the words are realized top-down. According to this classification, WP and IA are maximally different: WP is an inferential-realizational theory, while IA is a lexical-incremental theory.

Matthews' implementation of WP somewhat obscures his original proposal to assign fundamental status to words and paradigmatic relationships. Instead, the implementation looks more like an Item-and-Process approach, where rules add morphosyntactic properties and exponents onto stems within single words. His morphological transformations (Matthews 1991:193) are more in line with WP principles. Morphological transformations combine formal and semantic operations and operate over whole words. Thus, a morphological transformation may be used to derive one form in the paradigm (e.g. the genitive of a noun) from another form (e.g. the nominative).

The basic elements of an IA theory are a set of lexical entries (one for each morpheme, roots and grammatical morphemes alike) and a set of rules for putting them together. Lieber, working within a generative syntactic theory, suggests that morphemes are put together in binary branching trees. Compositionality is assumed as the basic principle in IA. The morphosyntactic properties of a word are constructed from morphosyntactic properties associated with the word's constituent morphemes.

In WP, on the other hand, compositionality is not a necessary fact. The meaning of the whole word licenses the exponents to be used, but there is no precondition that the meanings of the exponents have to combine to comprise the meaning of the whole. Compositionality may, indeed, emerge, but as a side effect rather than a central principle, or perhaps as an effective learning strategy. The whole itself may contribute meaning to

the meanings of the parts, or may override the meanings of the parts.

A principled distinction between WP and IA is in their views of the lexicon. In IA, the lexicon consists of entries for roots and affixes, and the goal of the theory is to produce economical descriptions, where the economy concerns the number of segments used (Bochner 1993). By contrast, WP approaches view the lexicon as containing the full word-forms necessary and sufficient for predicting the rest of the paradigms. Thus, instead of trying to minimize the number of symbols in a description, WP concentrates on the informational content of the stored exemplar words. The architecture of WP models assumes that a sufficient number of example paradigms is stored so that novel items can be formed by analogy.

Bochner (1993) is a fairly explicit and elaborated modern Word-and-Paradigm model, called Lexical Relatedness Morphology (LRM). LRM contains a set of rules for word formation. Rules are bi-directional relations between sets of related forms, not generative unidirectional rules as in IA or IP. The morpheme has no theoretical status within LRM (Bochner 1993:63). Bochner argues that paradigm-based morphology is in some sense more economical than morpheme-based models, if one uses a metric of how informative each lexical entry is, following (Jackendoff 1975). Similar measures of economy are discussed by Carstairs (1983) and Blevins (2004). According to Bochner, cases where form-to-meaning relations are not one-to-one contradict the symbol-counting metric, and even a few deviations are thus fatal to the morpheme-based theories from the point of view of economy of description. By contrast, the informational value metric and word-based theories do not have the same problems.

2.3.4 Modern WP Approaches

Most subsequent implementations of Matthews' and Robins' insights rely on realization, or 'spell-out' rules, to describe how morphosyntactic properties relate to full word forms. An example spell-out rule for the English plural is in (39). It associates the property 'plural' with a morphological rule that combines a stem and an affix.

(39) Realizational rule for 'plural' (cf. Matthews 1991, 175)

Plural		
/X/	→	/X+z/

The simple plural rule allows the a one-to-one correspondence between a morphosyntactic property and a unit of form. However, the format of the realization rules is designed in such a way that any number of properties can be associated with any number of form exponents, as Matthews demonstrates on the analysis of the Greek perfective formations (summarized in Blevins forthcoming).

The realization rules rely on abstracted representations of paradigm cells and must therefore provide some way to index the lexemes which they are describing. Thus, many realizational approaches take a form that Blevins (2006) has called "Stem-and-Paradigm," referring to the way words are constructed in these approaches from basic stems via realizational rules. These descriptions abstract away from paradigmatic comparisons in their representations, but the prose descriptions retain much of the implicational structure of paradigms described by Matthews. However, by contrast to the more traditional models, Stem-and-Paradigm models give a special ontological status to stems.

Aronoff (1994) provides arguments for the theoretical significance of stems as sound forms independent of meaning. He describes the formation of the Perfect and Future participles in Latin (table 2.9). Both participles have the suffix *-t*, which is not shared by the Present active infinitive. Moreover, the formation of the Future participle seems to be completely dependent on the Active participle, including idiosyncrasies in deriving the *-t* form from the infinitive, and any stem suppletion.

PRESENT ACTIVE INFINITIVE	PERFECT PARTICIPLE	FUTURE PARTICIPLE	GLOSS
laudā-re	*laudāt-*	*laudāt-ūr-*	'praise'
monē-re	*monit-*	*monit-ūr-*	'warn'
duce-re	*duct-*	*duct-ūr-*	'lead'
fer-re	*lat-*	*lat-ūr-*	'bear'

Table 2.9: Latin Perfect and Future participles (Aronoff 1994:32)

This pattern is problematic for morpheme-based approaches for several reasons. First, even though the Future participle seems to be morphologically derived from the Perfect participle, there is no sense in which its meaning is derived from that of the Perfect participle. Second, there are verbs that do not have a perfect participle but do still have a Future participle (e.g. *caleō, calitūr-* 'burn, be hot'). Finally, there are other Latin verb forms that appear to also be derived from the Perfect participle, which share even fewer

semantic or syntactic properties with it, such as the supine (Aronoff 1994:35). So, instead of deriving forms from the Future participle, Aronoff suggests that they share a common stem, the so-called 'third stem' (in addition to the present and perfect stems). The third stem is associated with each verb lexeme and does not carry any syntactic or semantic properties; its ontological status is merely as a morphological unit that can be used to create word forms.

Anderson (1992) and Stump (2001) both propose systems consisting of sets of stems and rules which operate on those stems to produce full word forms. In both systems, a derivational perspective supplants the traditional perspective base on exemplary paradigms. Stump's Paradigm Function Morphology additionally provides mechanisms by which realization rules for some paradigm cells may refer, or depend on, the rules, or output of rules, for other cells, thus accounting for 'rules of referral' described above.

Both Anderson and Stump are largely concerned with ordering rules in such a way that the resulting forms are correct. Anderson does this by putting rules in rule blocks and ordering blocks with respect to each other; rules within a block are intrinsically ordered by a specificity principle. Stump's approach modifies the interpretations of the rules themselves so that no intrinsic ordering applies across blocks. As I will argue in Chapter 3, both strategies are ultimately insufficient for the Georgian data because they are forced to make arbitrary decisions in ordering the rules. Moreover, the identification of a stem in Georgian is not by itself enough to predict the entire paradigm of a verb. Whole-word comparisons are key for paradigm predictability, making the reification of the stem an unnecessary step.

Thus I use the insights of WP, but not any of its technical implementations, and paradigms are discussed more informally. The current approach is paradigmatic in that contrasts between related words are crucial for determining word meanings (rather than assembling word meanings from the meanings of constituent morphemes), and novel items are formed by analogy to existing ones.

The theoretical framework presented in this dissertation draws on both Construction Grammar and Word-and-Paradigm theories without committing itself to any one specific implementation of those theories. I emphasize theoretical decisions that are crucial for analyzing the relevant data, and remain agnostic on many others.

2.4 Constructional Morphology: Proposed Framework

The framework of Constructional Morphology, as presented in this dissertation, is designed to represent non-compositional and extra-compositional morphosyntactic patterns as well as compositional ones. It brings together many of the approaches described above, adds an inflectional morphology component, and focuses on the syntax-morphology interface. Just as the theoretical frameworks that serve as its input, the Constructional Morphology framework is essentially exemplar-based and thus akin to phonological exemplar models like (Johnson 2005b, Johnson 2005a, Albright 2002, Wedel 2003, Skousen et al. 2002).

The goals in constructing the framework are as follows.

- Generality of solutions and representational formats.
- Seamless expression of sub-lexical, lexical, and phrase-level idiosyncrasies and regularities.
- Conformance to available psycholinguistic evidence and research.
- Potential for describing a wide range of languages and morphosyntactic phenomena.

2.4.1 Some Basic Definitions and Principles

The basic unit of organization in Constructional Morphology is a *morphological construction*, a set of constraints on form (phonological, morphological, or syntactic) and meaning (possibly including semantic and pragmatic constraints). Just as in Construction Grammar, each linguistic expression may instantiate one or more constructions. Constructions can be organized into inheritance hierarchies, described in section 2.4.4, and constitute the *constructicon* of a language, which structures the linguistic knowledge that speakers have. Constructions are inferred on the basis of multiple stored examples and provide a convenient locus for analogical extensions. However, their psychological reality as separate from their stored instantiations is questionable. In contrast to some constructional models' emphasis on maximal generalization, only those schematic constructions are posited for which there is empirical evidence of productivity.

Constructions may involve different-size chunks of linguistic structure, from a morpheme to a word to a syntactic phrase. Although there is nothing within CxG that prevents morpheme-level constructions, I argue based on the case studies in this dissertation that most constructions operate on the level of whole words or larger. Hay and Baayen (2005)

suggest, based on a review of available psycholinguistic evidence, that morphemes may exist in the mental lexicon, but only as units parasitic on words.

A given construction need not specify the full composition of a given linguistic expression; rather, the constraints expressed by a construction may under-specify the form or the meaning of a linguistic expression. That is, constructions can be schematic. The constructions introduced in this dissertation will take the form of *templates*, imposing constraints on morphological or syntactic content of linguistic expressions.

In general, any set of form-meaning constraints that cannot be derived compositionally from the form-meaning constraints on its constituent elements is considered a construction. Given the prevailing view of semantic compositionality, this would seem to imply that very few things can be constructions. However, as argued in the CxG literature (Goldberg 1995, Fillmore et al. 1988), even argument structures impose some constraints not predictable from their component parts. The case studies in this dissertation suggest an even wider range of non-compositional constraints on grammar, including aspects of inflectional morphology.

Constructional approaches do not currently provide a uniform format for representing constructions. Some papers use box notation (Fillmore 1988), while others use feature structures (Ginzburg and Sag 2000). For ease of readability by the general linguistic audience, I will adapt an HPSG-like feature structure notation, greatly simplified so as to focus on the important aspects of the constructions.

The key components of a constructional representation are in (40). These are:

- The top-level sign type is "Construction."
- The relevant groups of features include SEM(antics), SYN(tax), MORPH(ology), and PHON(ology).
- Not every sign must have all of the above groups specified.
- The SYN values consist simply of a flat list of syntactic dependents, as will be seen in later representations.
- As suggested in (Kay and Fillmore 1999), the construction may impose constraints on (i.e. have access to) the structures inside its constituents. Both Construction Grammar and HPSG approaches make a point of emphasizing this requirement, based on the various syntactic constructions examined. As will become clear from the Georgian-specific morphological constructions in the following chapters, it may be needed for the morphology as well.

(40)
$$\begin{bmatrix} \textit{construction} & \\ \text{SEM} & \langle \ldots \rangle \\ \text{SYN} & \langle \ldots \rangle \\ \text{MORPH} & \langle \ldots \rangle \\ \text{PHON} & \langle \ldots \rangle \end{bmatrix}$$

Example (41) shows the representation for a fixed syntactic construction, "Kick the bucket."

(41)
$$\begin{bmatrix} \textit{construction} & \text{"Kick the bucket"} & \\ \text{SEM} & \begin{bmatrix} \textit{frame} & \text{"die"} \end{bmatrix} & \\ \text{SYN} & \begin{bmatrix} \boxed{1} \begin{bmatrix} \textit{construction-verb} & \\ \text{SYN} & \begin{bmatrix} \text{VERB} \end{bmatrix} \\ \text{PHON} & \textit{kick} \end{bmatrix}, & \boxed{2} \begin{bmatrix} \textit{construction-NP} & \\ \text{SYN} & \begin{bmatrix} \text{NOUN} \end{bmatrix} \\ \text{PHON} & \textit{the bucket} \end{bmatrix} \end{bmatrix} \\ \text{MORPH} & \langle \rangle & \\ \text{PHON} & \begin{bmatrix} \boxed{1}.\text{PHON} + \boxed{2}.\text{PHON} \end{bmatrix} & \end{bmatrix}$$

The important parts of the representation are:

- The SYN(tax) part looks like a regular syntactic feature structure.
- The SEM(antics) part need not be in a direct relationship to the syntax. In this case, it is not derived from the syntax at all.
- The construction essentially represents instructions for assembling the verb *kick* with the NP *the bucket*, but the overall meaning is different from what one would expect from a normal combination of the two parts. Thus, this construction could be described as a subcase of a regular NP-VP construction, with exceptional meaning. Describing it as a subcase has the advantage of not having to specify again the constraints on subject agreement, tense marking, etc.
- The semantics of a construction could, in principle, be represented in any way. CxG, however, is intimately connected to the theory of Frame Semantics (Fillmore 1982),

and I will use frames where possible for lexical meanings. A *frame* is essentially a scenario that provides background for the understanding of lexical meanings. Each frame specifies expected participants and semantic relations between them. So, the frame in which the verb *buy* is understood involves such participants as a seller and a buyer, and a scenario in which the seller gives the buyer goods in exchange for money.

2.4.2 Constructions at the Sub-Word Level: Inflection

As mentioned above, a construction can designate any linguistic expression or schema whose meaning cannot be inferred from the meanings of its component parts. I would like to suggest that inflectional morphology may also belong to this category. I will argue in Chapter 3 that elements of Georgian inflectional morphology are quite non-compositional. As a simpler example for now, let us examine case formation in Russian.

Russian nouns belong to one of three declension classes, one of which is subdivided into two groups (Table 2.10, cf. Timberlake 2004). The declensional classes are largely associated with additional semantic features such as gender, as indicated in the table, although this association is not without exception.

Decl.	Ia (masc.)		Ib (neuter)		II (fem.)		III (fem.)	
	SG	PL	SG	PL	SG	PL	SG	PL
NOM	*stol*	*stol-y*	*bljud-o*	*bljud-a*	*vilk-a*	*vilk-i*	*kost'*	*kost'-i*
ACC	*stol-a*	*stol-y*	*bljud-a*	*bljud-a*	*vilk-u*	*vilk-i*	*kost'*	*kost'-i*
GEN	*stol-a*	*stol-ov*	*bljud-a*	*bljud*	*vilk-i*	*vilok*	*kost'-i*	*kost'-ej*
DAT	*stol-u*	*stol-am*	*bljud-u*	*bljud-am*	*vilk-e*	*vilk-am*	*kost'-i*	*kost'-am*
INST	*stol-om*	*stol-ami*	*bljud-om*	*bljud-ami*	*vilk-oj*	*vilk-ami*	*kost'-u*	*kost'-ami*
LOC	*stol-e*	*stol-ax*	*bljud-e*	*bljud-ax*	*vilk-e*	*vilk-ax*	*kost'-i*	*kost'-ax*
Gloss	'table'		'dish'		'fork'		'bone'	

Table 2.10: Declension Classes of Russian Nouns

A look at the case endings reveals that they cannot have individual meanings in isolation from the declension classes with which they are used (unless one proliferates homophonous affixes). For example *-a* designates the Accusative and Genitive singular of class I, the Nominative and Accusative plural of class Ia, and the Nominative singular of class II. Thus, the suffix *-a* is just a formative without a uniquely associated meaning. It

participates in various case-marking constructions, but in and of itself it does not carry meaning[4].

The only way to infer the case endings for different declension classes is by comparing different case forms of a given word—that is, by looking at paradigmatic contrasts as suggested by Word-and-Paradigm approaches described above. Once morphological structure is inferred this way, paradigms for many words most likely stay in a speaker's mental lexicon. This must obviously be true for irregular forms, but there is evidence that even regular, productive inflected forms are stored if they are sufficiently frequent (Bybee 2001). Lexical classes can be recognized on the basis of paradigmatic contrasts; typically, one or more inflected forms of a word are needed to infer its lexical class and, therefore, the rest of its paradigm.

On the other hand, once morphological patterns are inferred on the basis of example paradigms, they may acquire a more schematic status (although probably still grounded in the actual example words), particularly in the case of frequent patterns. Exemplary patterns may then analogically extend to novel items, similarly to the process described by Blevins (2006). A novel form may be associated with an inflectional class if a sufficiently informative form of it is known (such as the nominative singular, sometimes in combination with gender, in Russian); or, if the novel form is uninformative, it may be associated to the most frequent and/or the most productive inflectional pattern.

For Russian, the case endings must be specified in conjunction with the morphosyntactic property "case" and the declension class of a particular noun. Constructionally, this can be represented as in (42).

[4]There is some evidence from priming studies that the idea of a meaningless 'morphological formative' may be applicable even to cases of accidental homophony, in affixes as well as roots (like 'sew' and 'sow' in English). The mention of a homophonous root in a disambiguating context nevertheless primes other words in which such roots may participate (Zwitserlood 1996).

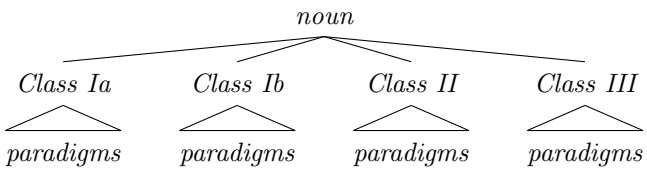
(42)
$$\begin{bmatrix} \textit{cxn-GenPL-Ia} & \\ \text{SEM} & \begin{bmatrix}\text{GENITIVE MEANING}\end{bmatrix} \\ \text{SYN} & \begin{bmatrix}\textit{cxn-noun}\end{bmatrix} \\ \text{MORPH} & \begin{bmatrix} \text{CLASS} & \text{IA} \\ \text{CASE} & \text{GEN} \\ \text{NUMBER} & \text{PL} \end{bmatrix} \\ \text{PHON} & \begin{bmatrix}\text{X}+/ov/\end{bmatrix} \end{bmatrix}$$

The 'X' in the construction in (42) represents an abstracted formal stem to which the formal affix is attached, as in (Matthews 1991). Crucially, however, neither the stem nor the affix themselves have meanings individually associated with them. The representation in (42) is only possible as an abstraction over multiple examples of class Ia genitives. The "genitive meaning" and other case meanings could be represented as radial categories with prototypical meanings and construction-specific extensions (Janda and Clancy 2002).

Also, one should keep in mind that the ontological status of schematic constructions as in 42 is questionable. While it is clear that whole words are stored, and that more general patterns may be extracted from them, there is no clear evidence that the abstract schemas are stored separately from the examples that gave rise to them (see section 2.5). While the abstract schemas are a convenient way for the linguist to represent general patterns, the generalizations are likely stored in a distributed fashion. I will assume here that such representations are a notational convenience, and thus the particular way in which the phonological constraints are represented is not crucial.

I am assuming that declension classes can be represented by a lexical inheritance hierarchy, as is typical in HPSG accounts (43). The crucial difference is that the inheritance hierarchy is only an abstraction over existing words and constructions.

(43) Russian inflectional classes



A paradigm is essentially a structured set of relations between individual words (or fully specified constructions). A case paradigm for the word *stol* can be represented as in Figure 2.1. A more abstract representation for the paradigm of class Ia nouns is in Figure 2.2.

These constructions can be used for analogical extensions to form novel items. Generally, the presence of an underspecified construction like NomSg-Ia indicates that there is sufficient evidence for a generalization, and such a generalization can be used for analogical extensions. While it remains an empirical question, it is likely that speakers retain some combination of constructions and stored paradigms.

In this system, rules of referral can be represented fairly straightforwardly, by reference within one construction to the construction for the appropriate paradigm cell. For instance, Russian class Ia nouns have identical nominative and accusative forms. This relations can be represented as in (44), where X is the lexeme index (not the stem). The form and meaning elements of different constructions can be referred to independently of each other.

(44)
$$\left[\begin{array}{ll} \textit{cxn-AccSg-Ia-X} & \\ \text{MORPH} & \left[\begin{array}{ll} \text{CASE} & \text{ACC} \\ \text{NUM} & \text{SG} \end{array}\right] \\ \text{PHON} & \text{CXN-NOMSG-IA-X.PHON} \end{array}\right]$$

Another influence of WP approaches on the proposed framework is that the morphological constructions presented here operate on whole words. Of course, not every construction licenses a fully-specified word; rather, individual constructions may impose constraints on different parts of a word, but with the understanding that those parts are embedded in a larger whole. Thus meaning is never associated to individual morphs. Roots are the most obvious morphological elements with a claim to individual meaning; however, in an exemplar-based system, roots are merely abstractions over sets of related words Kuryłowicz (1949b, 1949a), and what could be considered their individual 'meaning' is distributed across words that instantiate it. This model works particularly well for languages where all roots are bound (arguably, Georgian is such a language), but it also works for languages like English where most roots can occur as full words.

Neither roots nor affixes are listed separately in the lexicon. The form part of morphological constructions represents a morphological template and, depending on how

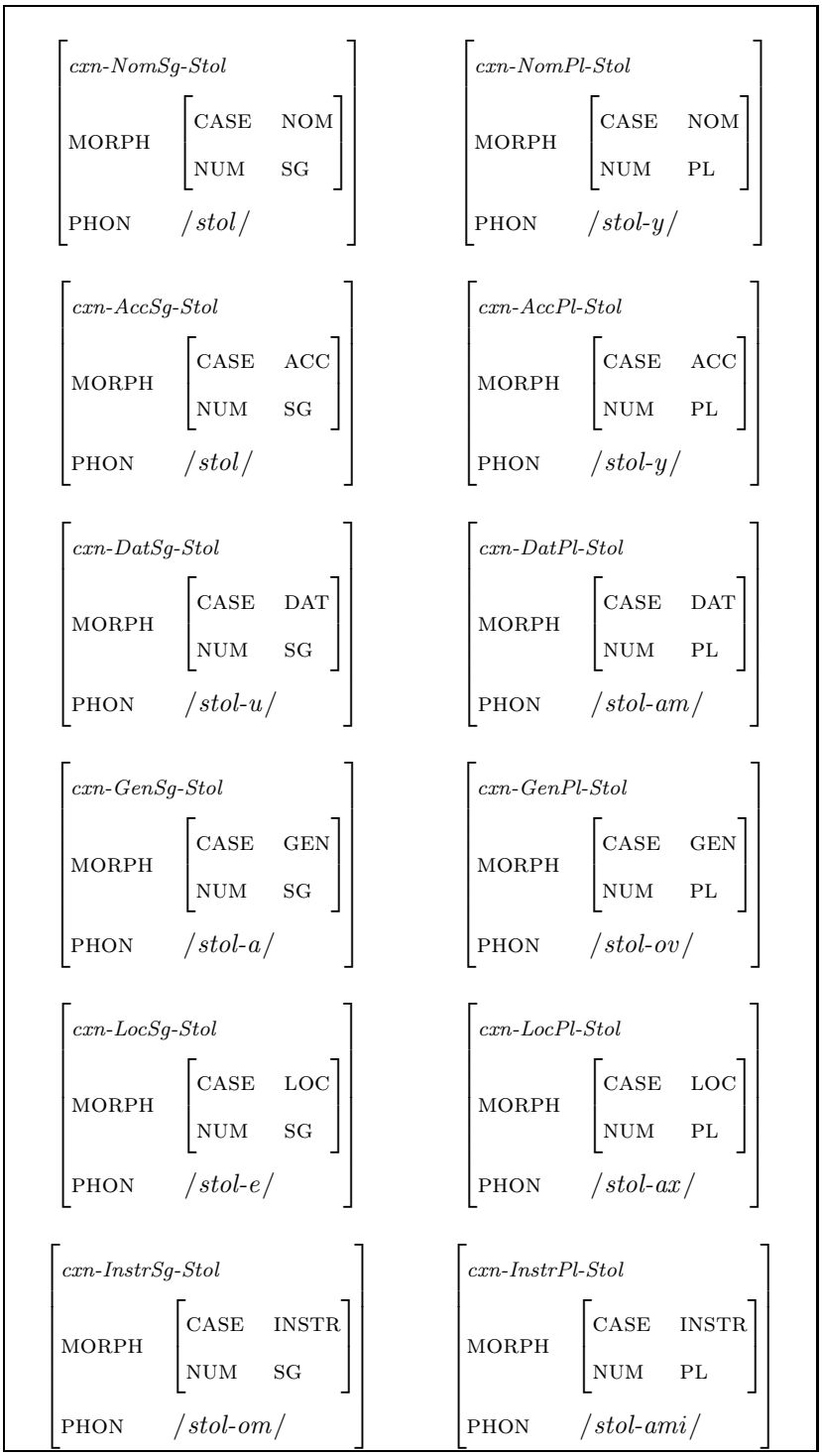


Figure 2.1: Constructional Representation of a paradigm for *stol*

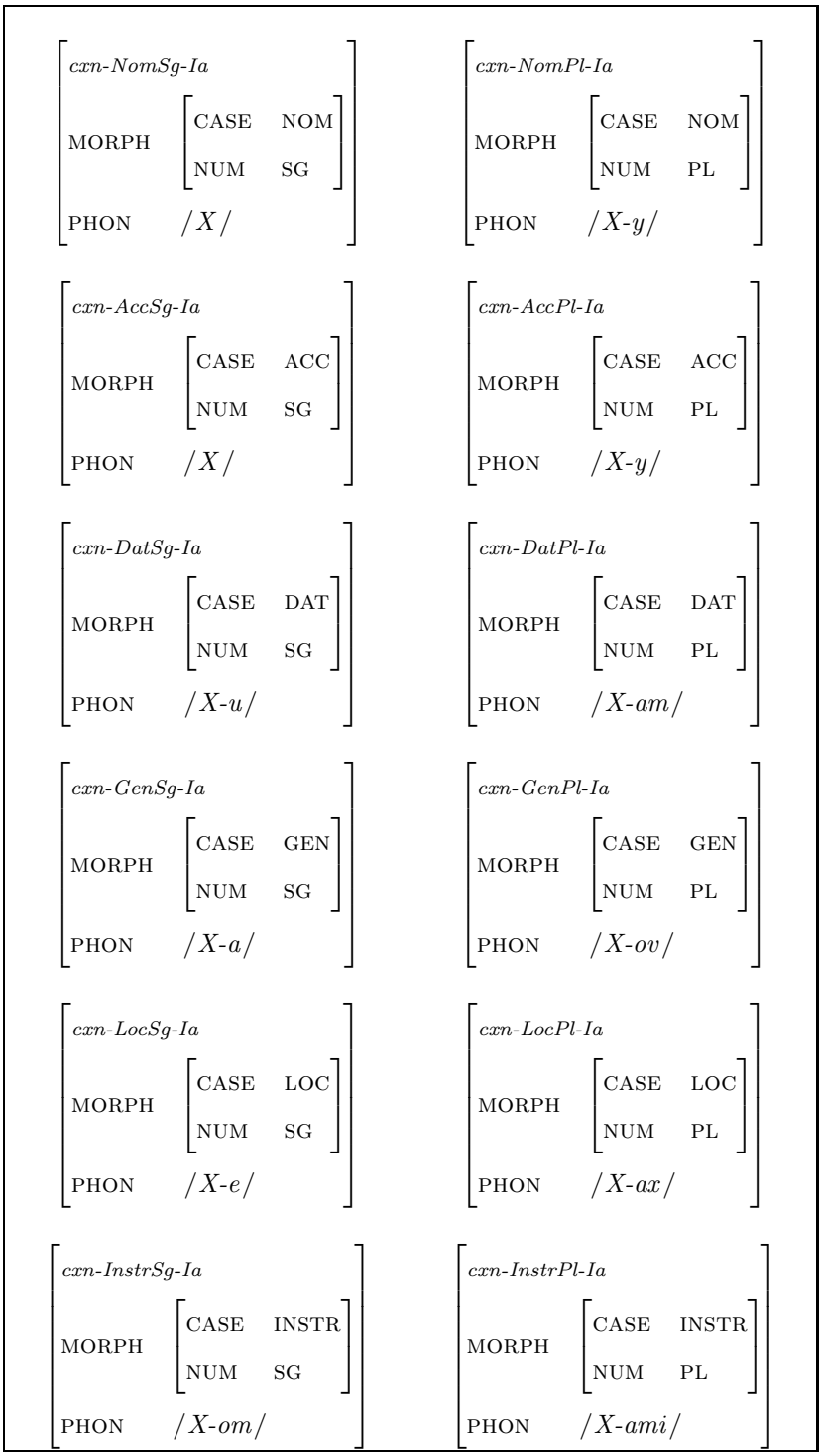


Figure 2.2: Constructional Representation of a paradigm for class Ia nouns

abstract a given construction is, it can be more or less specified. The morphological templates specify constraints on the output rather than on the derivational process that goes from constituent morphemes to complete words. They are thus similar to product-oriented schemas (Bybee 2001:126), which have been shown to be very important in structuring the mental lexicon of a speaker (Burzio 1996, Burzio 2002).

Constructional patterns provide a reasonable mechanism for how novel words are inflected, once a word's declension class is determined. Essentially, they represent schemas that retain the productive aspects of morphological formation while abstracting away from individual word forms.

On the other hand, constructions can also express fixed, non-productive, or irregular patterns, as is emphasized in the CxG literature. The formation of the genitive plural in Russian is sometimes irregular. For example, the word *soldat* 'soldier', from declension class Ia, has the genitive plural *soldat* instead of the expected **soldat-ov*. Constructionally, this pattern could be represented as in (45), with the stem fully specified. The interaction of this kind of construction and the more regular constructions is described more fully in section 2.4.4.

This is essentially equivalent to listing the exceptional word, the way a morpheme-based approach would do. The difference is that many words will be listed, not just the irregular ones. Whether or not a word is stored is determined by its token frequency (as suggested by Bybee (2001) and others) rather than by whether or not it instantiates a regular pattern.

(45)
$$\left[\begin{array}{ll} \textit{cxn-GenPL-soldat} & \\ \textsc{sem} & \left[\textsc{Soldier} + \textsc{Genitive meaning}\right] \\ \textsc{syn} & \left[\textit{cxn-noun}\right] \\ \textsc{morph} & \left[\begin{array}{ll} \textsc{class} & \textsc{Ia} \\ \textsc{case} & \textsc{gen} \\ \textsc{number} & \textsc{pl} \end{array}\right] \\ \textsc{phon} & \left[/\textit{soldat}/\right] \end{array}\right]$$

The types of regular patterns and irregularities discussed so far could conceivably be accommodated in a morpheme-based framework, as well. The power of a construction-

based approach becomes more evident when we examine partial regularities and sub-regularities that extend across established lexical classes. Nominal stress patterns in Russian provide a good example.

There are three main stress patterns, which divide into six sub-patterns (Timberlake 2004) that cut across declension classes, distinguished by where the stress falls in singular and plural forms.

- (a) Stress on the root in both singular and plural, e.g. *nedélja – nedéli* 'week'
- (b) Stress on the ending in both singular and plural, e.g. *stól – stolý* 'table'
- (c) Stress on the ending in the singular and the root in the plural, e.g. *oknó – ókna* 'window'
- (d) Stress on the root in the singular and the ending in the plural, e.g. *slóvo – slová* 'word'
- (e) Stress on the root in the singular and plural, but retracted in the Nom and Acc plural, e.g. *zúb – zúby – zubám* 'tooth'
- (f) Stress on the ending in the singular and the ending in plural, but retracted in the Nom and Acc plural, e.g. *kón' – kóni – konjám* 'horse'

These stress patterns are distributed throughout the declension classes, and it is impossible to predict from the declension class which stress pattern a noun will take, or vice versa. However, the number of stress patterns is small, and within a given stress pattern all stresses in a paradigm are predictable. Thus, stress patterns present a set of patterns that are orthogonal to declension class but can be abstracted from the same example nouns as the declension classes. Usually, a comparison between a singular form and a plural form are sufficient to predict the stress pattern (patterns (e) and (f) need two plural forms, but they are less frequent and unproductive). A morpheme-based approach would have to index each noun for both its declension class (in order to assign correct inflectional endings) and stress pattern, multiplying the number of categories of noun lexemes. An abstractive, word-based approach could abstract a multidimensional set of patterns from stored examples without multiplying diacritic features. The resulting network of patterns can be represented as in Figure 2.3.

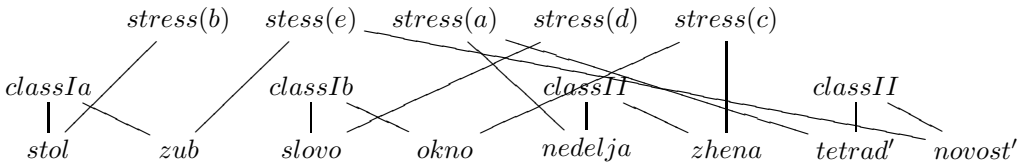


Figure 2.3: Nominal Declension and Stress Patterns in Russian.

2.4.3 Morphosyntactic Constructions

Subject-Verb Agreement in Russian

Morphosyntactic constructions represent a level of abstraction somewhere between the purely morphological and the purely syntactic expressions. The difficulty in representing such constructions is that the interactions between syntax and morphology must be clearly described and delineated.

As an example of morphosyntactic constructions, let us examine subject-verb agreement in Russian. Verbs that take nominative subjects must agree with them in several features; constructions with non-nominative subjects follow other agreement principles which will not be discussed here. Verbs in the present or future tense must agree with their subject in person and number. The future-tense verbal paradigm for *prinesti* 'to bring (perfective)' is represented in Table 2.11.

	SG	PL
1ST PERSON	*prines-u*	*prines-ëm*
2ND PERSON	*prines-ëš*	*prines-ëte*
3RD PERSON	*prines-ët*	*prines-ut*

Table 2.11: Russian *prinesti* 'to bring (perfective)' in the future

In the past tense, verbs do not inflect for person, but for gender instead. In Common Slavic, the past tense was expressed periphrastically by an adjective-like participle and an auxiliary form of the verb 'to be' (Schenker 1995:148). The auxiliary was lost in Russian, leaving just the participle which declines for gender. The past-tense paradigm for the verb 'to bring (perfective)' is in Table 2.12.

This discrepancy in agreement features is the result of a historical accident, and

	SG	PL
MASCULINE	*prinës*	*prinesl-i*
FEMININE	*prinesl-a*	
NEUTER	*prinesl-o*	

Table 2.12: Russian *prinesti* 'to bring (perfective)' in the past

no principled combination of verbal features would predict it. Rather, agreement in the present/future and the past is guided by two separate sets of constructions, which are essentially stipulations and must be learned by speakers.

Representing agreement requires a number of semantic, syntactic, morphological, and phonological features. The person and gender of the subject are semantic—in cases where semantic gender of the subject is in conflict with morphological gender, the verb agrees with semantic gender (46). The case of the subject is a syntactic feature. The agreement features marked on the verb are morphological, and are expressed phonologically.

(46) *Vrach prinësla knigu.*
doctor brought.FEM book.ACC
'The doctor brought a book.'
(*Vrach* is morphologically masculine but here semantically feminine)

A constructional representation of the present/future tense agreement is in (47); the past tense agreement construction is in (48).

(47)
$$\begin{bmatrix} \textit{cxn-Subj-Verb-Future} & & & \\ \text{SEM} & \begin{bmatrix} \textit{frame "Actor-Action"} & \\ \text{ACTOR} & \boxed{1}.\text{SEM} \\ \text{ACTION} & \boxed{2}.\text{SEM} \end{bmatrix} & & \\ \text{SYN} & \begin{bmatrix} \boxed{1} \begin{bmatrix} \textit{cxn-noun} & & \\ \text{MORPH} & \begin{bmatrix} \text{CASE} & \text{NOM} \\ \text{PERSON} & \boxed{3} \\ \text{NUMBER} & \boxed{4} \end{bmatrix} \end{bmatrix}, \boxed{2} \begin{bmatrix} \textit{cxn-verb} & & \\ \text{MORPH} & \begin{bmatrix} \text{TENSE} & \text{FUTURE} \\ \text{PERSON} & \boxed{3} \\ \text{NUMBER} & \boxed{4} \end{bmatrix} \end{bmatrix} \end{bmatrix} & \end{bmatrix}$$

(48)
$$\begin{bmatrix} \textit{cxn-Subj-Verb-Past} & \\ \text{SEM} & \begin{bmatrix} \textit{frame ``Actor-Action''} & \\ \text{ACTOR} & \boxed{1}.\text{SEM} \\ \text{ACTION} & \boxed{2}.\text{SEM} \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \boxed{1} \begin{bmatrix} \textit{cxn-noun} & \\ \text{MORPH} & \begin{bmatrix} \text{CASE} & \text{NOM} \\ \text{GENDER} & \boxed{3} \\ \text{NUMBER} & \boxed{4} \end{bmatrix} \end{bmatrix}, \boxed{2} \begin{bmatrix} \textit{cxn-verb} & \\ \text{MORPH} & \begin{bmatrix} \text{TENSE} & \text{PAST} \\ \text{GENDER} & \boxed{3} \\ \text{NUMBER} & \boxed{4} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The key component of this representation is that a construction places constraints on both the verb and its subject simultaneously, and any of the features may be affected—morphological, syntactic, or semantic.

Valency

The representation of Georgian version requires a set of fairly complex morphosyntactic constructions. In addition to cross-referencing features of the verb and its arguments, version also requires a representation of valency and valency alternations. Chapter 4 presents a detailed analysis and a worked-out representation. For now, I will introduce the general mechanism for representing valency.

In lexicalist grammars, valency restrictions are typically represented as properties of individual verbs. The more general patterns like transitivity are usually assigned to more abstract lexical types from which individual verbs inherit. By contrast, constructional approaches tend to view valency as a restriction on a whole *construction*, which comprises the verb and its arguments. In keeping with a constructional view, (49) illustrates the representation of a transitive construction without overt morphology, such as *take a book* in English.

(49)
$$\begin{bmatrix} \textit{cxn-Verb-Obj-NoMorph} & \\ \text{SEM} & \begin{bmatrix} \textit{frame ``Action-Patient''} \\ \text{ACTION} \quad \boxed{1}\text{.SEM} \\ \text{PATIENT} \quad \boxed{2}\text{.SEM} \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \boxed{1}\begin{bmatrix}\textit{cxn-verb}\end{bmatrix}, \boxed{2}\begin{bmatrix}\textit{cxn-noun}\end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Adding overt morphology, (50) presents a possible transitive construction in Russian where the object must be in the accusative case.

(50)
$$\begin{bmatrix} \textit{cxn-Verb-Obj-Acc} & \\ \text{SEM} & \begin{bmatrix} \textit{frame ``Action-Patient''} \\ \text{ACTION} \quad \boxed{1}\text{.SEM} \\ \text{PATIENT} \quad \boxed{2}\text{.SEM} \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \boxed{1}\begin{bmatrix}\textit{cxn-verb}\end{bmatrix}, \boxed{2}\begin{bmatrix} \textit{cxn-noun} \\ \text{MORPH} \quad \begin{bmatrix}\text{CASE} \quad \text{ACC}\end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

2.4.4 Putting It Together: Constructs and Constructional Hierarchies

As suggested by Kay and Fillmore (1999), constructions are schematic representations of linguistic patterns. A construction can be instantiated by a *construct*, a real linguistic expression. Each expression may simultaneously instantiate more than one construction. In such cases, the properties of the construct are determined by unifying the various constructions. Another way of thinking about this is that each stored expression can participate in the abstraction of multiple unrelated patterns, in a multidimensional lattice of such patterns.

To demonstrate how this works, let us examine a simple transitive sentence in Russian (51).

(51) *Ja prinesla* *knigu.*
I brought.PAST.FEM book.ACC
'I (fem.) brought a book.'

Many of the constructions instantiated by the expression in (51) were described above: the accusative construction has similar components to the genitive construction in (42), the transitive verb-object construction (50), and the verb-subject agreement construction (48). The result of unifying all of these is in Figure 2.4. The diagrams are greatly abbreviated for the sake of readability; in particular, lexical meanings are not included.

Constructions can capture patterns of various specificity. In order to describe the relationships between similar constructions at various degrees of abstraction, they are organized into hierarchies. More specific constructions inherit properties from the more general constructions. By default, all properties of constructions higher up in the hierarchy are passed on to those lower down. However, following Goldberg (1995) and Kay and Fillmore (1999), the more specific constructions may have properties that conflict with those of their parents, and the more specific properties override the more general ones.

Taking an abstractive view, one must allow for the formation of generalizations despite the presence of counterexamples. Riehemann (2001) cites the difficulty of doing this in language acquisition as the basis of choosing monotonic inheritance rather than default inheritance. However, the fact that children start with lexical patterns, go through an overgeneralization stage, and finally settle down on the irregular patterns suggests that such learning is, indeed, possible. (cf. Bowerman 1982, Kuczaj 1977, Hoff 2001). On the other hand, the two types of representations (default and monotonic) seem to be isomorphic, but the default version is intuitively better for contrasting regular and irregular patterns.

To take the example of Russian genitive plurals, an inheritance hierarchy would look something like Figure 2.5. There is a general "genitive plural" construction, which is parent to a "genitive plural class Ia construction". The latter is the default pattern for the formation of genitive plurals for nouns like *stol*∼*stolov* 'table' and *slon*∼*slonov* 'elephant'[5]. An exceptional noun like *soldat*∼*soldat* 'soldier' uses a more specific construction, which inherits from the "genitive plural class Ia" all of its semantic and syntactic properties, but overrides the constraints on the phonology. The overriding relation is indicated by a dashed line in the hierarchy. This analysis borrows from that of Hippisley (1999), which also relies on a default inheritance hierarchy with possible overrides. The paradigms serve as the basis for inferral of other inflectional patterns in parallel, creating a network of abstracted

[5]There are two productive patterns of forming the genitive plural of class Ia nouns; the choice is determined by whether or not the stem-final consonant is determined by the stem-final consonant: *stol*∼*stol-ov* 'table', *nož*∼*nož-ej* 'knife'

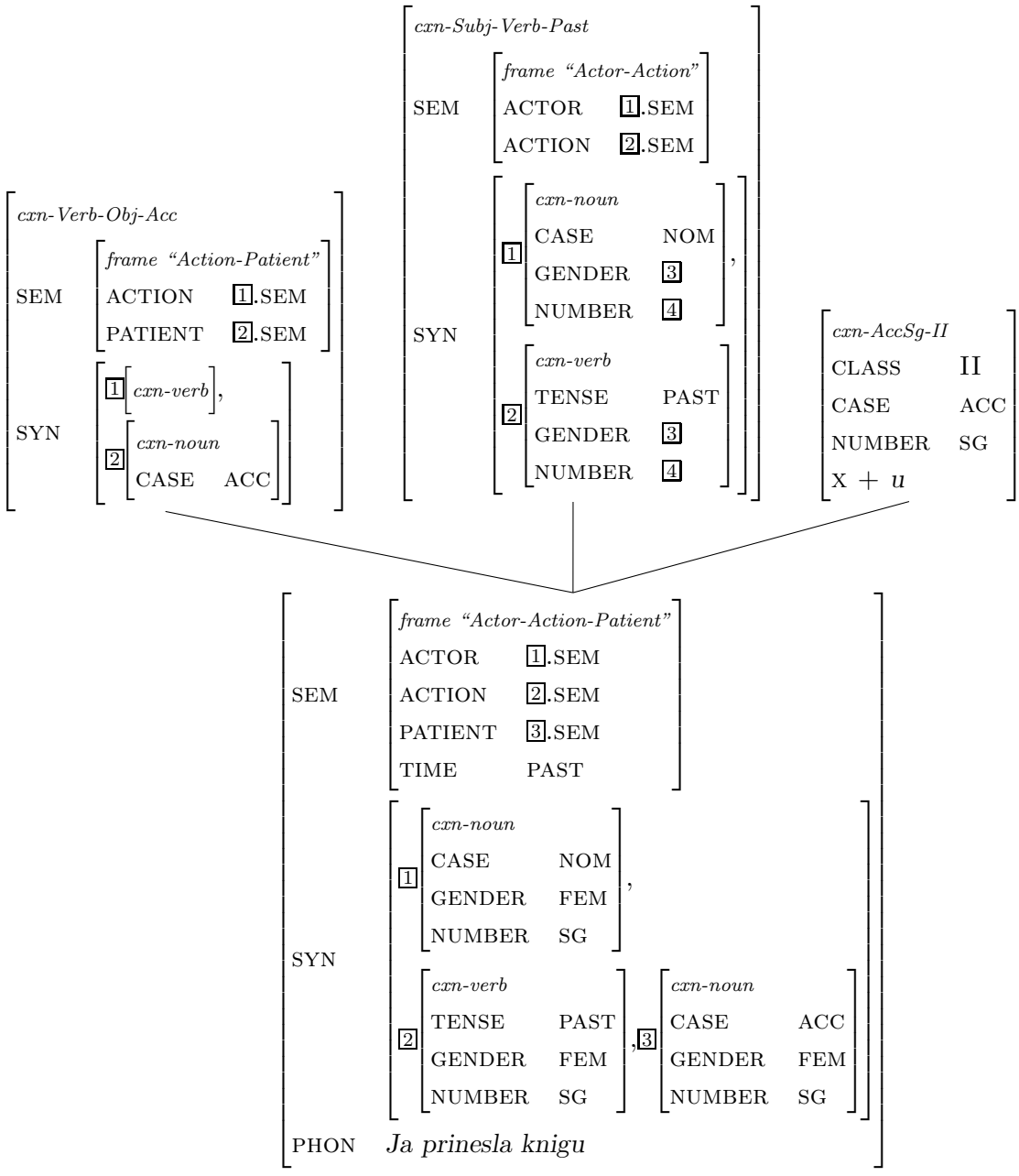


Figure 2.4: Integrating case and agreement constructions

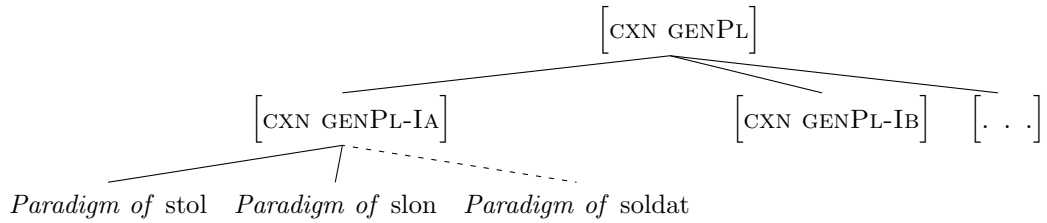


Figure 2.5: Russian case construction hierarchy

patterns. The network is not represented here for the sake of readability, but it should look similar to the informal network in Figure 2.3.

Another case of a morphological default pattern with overrides is described by Hyman (2002) for Bantu affix ordering. Hyman describes several variable orderings of the Causative, Applicative, Reciprocal, and Passive suffixes in Bantu. Analyses based on the Mirror Principle (Baker 1985) would predict that affix ordering should be determined by compositional semantic principles: the order of application of semantic operations should correspond to the order of morphological attachment to the root. However, an examination of the Bantu data reveals that there are many deviations from the semantically predicted order. Instead, Hyman argues on comparative and historical evidence that there is default *morphological* template that is independent of semantic principles (52).

(52) Bantu "default" template C-A-R-P (Hyman 2002)

CAUS		APP		REC		PASS	
-ic-	>	-id-	>	-an-	>	-u-	Proto-Bantu
-is-	>	-il-	>	-an-	>	-w-	Shona
-ih-	>	-il-	>	-an-	>	-iw-	Makua
-its-	>	-il-	>	-an-	>	-idw-	Chichewa

If CARP is considered the default ordering, then all of the attested pairwise deviations turn out to be semantically (Mirror) motivated. Hyman accounts for the ordering facts within Optimality Theory using a conjunction of the templatic ordering constraint and the Mirror Principle constraint as compared to the templatic constraint alone. However, this account does not fully capture the generalization that the morphological template is the *default* ordering, and is violated only when discourse and semantic factors are prominent. In a constructional hierarchy, this generalization can be naturally accommodated by viewing the morphological order as default, with the deviations as semantically- and

pragmatically-driven overrides of the default. Since constructions can impose constraints on any level of the linguistic expression, the fact that one ordering is morphological and the other is semantic is not an issue.

The inheritance relations between constructions are very similar to the instantiation relations between constructions and constructs. At this point, I am agnostic as to whether the two types of relations are in fact the same. There seem to be several differences between them, but these could conceivably be reconciled even if instantiation were a kind of inheritance.

First, linguistic expressions most often make use of multiple constructions of different kinds. If instantiation were a kind of inheritance, we would need mechanisms for multiple inheritance at the lowest levels of the construction hierarchy. Goldberg (1995) presents arguments in favor of multiple inheritance of constructions. Her examples, however, involve specific linguistic expressions, and it is not clear whether or not multiple inheritance is needed at higher, more abstract levels of the construction hierarchy.

Second, the constructions represented above are fixed patterns, presumably stored in speakers' memories, whereas concrete linguistic expressions can be made up on the fly. One has to keep in mind, of course, that specific linguistic expressions are the basis for forming constructional generalizations during language acquisition. The question of when a linguistic expression becomes a construction seems to underlie the decision above.

Tomasello (2003) suggests that all linguistic expressions start out as memorized, and exposure to multiple instances of the same pattern make possible the formation of generalizations. Thus it seems that children's earliest expressions are constructions in Goldberg's sense: since children do not know the more general patterns, the meanings of these simple constructions are not predictable from the meanings of their parts (or, perhaps, they are not even divisible into parts). As suggested by Bybee (2001), these more specific constructions may not be discarded even when larger generalizations are formed. So until the relationships between compositional predictability, productivity, frequency, and memory storage are determined empirically, it seems senseless to theorize on the difference between instantiation and inheritance.

It seems that in either case, the description of specific linguistic expressions must make use of some sort of constraint satisfaction mechanism. However, the exact nature of the relation between linguistic expressions and more fixed constructions might not be theoretically significant.

2.5 Productivity and Well-Formedness

2.5.1 Psycholinguistic Evidence and Analogy

A key notion behind both Word-and-Paradigm and constructional models is that they are *exemplar-based*. Patterns are abstracted away from concrete examples, and the original word forms of phrases that gave rise to the patterns do not disappear once a generality has been detected. The knowledge of a pattern is thus distributed across many specific examples. Kay (2002) hints at this notion for Construction Grammar:

> A formal difference between C[x]G and (I think) both LFG and HPSG is that C[x]G admits only actual words, phrases and sentences (constructs) as elements of the modeling domain. In HPSG terms, this would roughly amount to saying that the only modeling domain objects are signs. Abstractions over actual linguistic objects (words, phrases, sentences, i.e., constructs) are not considered to be part of the modeling domain in C[x]G. (Kay 2002:17)

And for Word-and-Paradigm, the fundamental status of example paradigms is at the core of the model, as confirmed by Blevins (2006) below[6].

> Taken together, exemplary paradigms and principal parts provide the information required to deduce new forms. The special status of an exemplary paradigm lies in the fact that it exhibits general patterns of inflection. A set of principal parts contributes item-specific word forms, which establish a link to the exemplary pattern that provides the model for the inflection of an item. Matching the principal parts of an item against the cells of an exemplary paradigm establishes a correspondence between principal parts and their counterparts in the exemplary paradigm. New forms of an item are deduced by extending this correspondence analogically to other cells. (Blevins 2006)

Much recent work in psycholinguistics and language acquisition provides support for exemplar-based, constructional description of linguistic generalizations. Acquisition has been shown to be item-based, where lexical constructions are gradually extended to include several words of one type (Dodson and Tomasello 1998, Tomasello 2003). Children start with single-word constructions, slowly learning two-word fixed constructions, and are very conservative about extending generalizations to novel words. There appears to be no evidence for learning large-scale generalizations independently of the lexical items that

[6]Morpurgo Davies (1978, 1992) discusses the origins of notions of analogy in the work of the Neogrammarians, suggesting that it was a central idea to the architecture of the grammar.

exemplify them, even when learning argument-structure generalizations (Goldberg et al. 2004)[7].

This is especially true of verb learning. Dodson and Tomasello's concept of 'verb islands' refers to the fact that children start out by pairing specific verbs with specific nouns, so that the same verb is always associated with the same noun. Different verbs are used with different prototypical nouns, regardless of the syntactic similarities between verbs. Slowly over time, children seem to abstract away verb frames and increase the range of nouns with which a certain verb is used. However, even with different nouns, a single verb is usually associated with only one valence and word-order pattern. This suggests that linguistic learning, at least at the early stages, is exemplar-based and not rule-based.

Some recent experiments on children's acquisition of novel argument structure patterns shows that patterns are acquired (and generalized to novel verbs) significantly better if the input is "skewed," i.e. one verb is responsible for a disproportionately large number of instances of the new construction, rather than several verbs being equally frequent in the novel construction (Casenhiser and Goldberg 2005). This is parallel to what happens in actual acquisition, where, e.g, the verb 'give' is in the majority of ditransitive constructions heard by children (Cameron-Faulkner et al. 2003). This suggests very strongly analogy from a single lexical item to a novel lexical item, rather than the application of a general scheme. The same appears to be true for generalizing non-linguistic patterns such as dot formations.

The notion of analogy as the mechanism for productivity is somewhat more problematic. While it is clear that the human brain is capable of analogical extensions, it is not yet known exactly how analogy operates. One study has demonstrated that a four-part analogy (e.g. A:B = C:x) works when there is structural similarity between the two patterns (Gentner and Markman 1997); structural similarity was shown to be more important than lexical or semantic similarity between the corresponding elements of the two patterns. For instance, in an analogy A:B=B:x, the structural similarity between A and B was found to be more important than the 'lexical' similarity between B in the first position and B in the second position. Analogy, is pervasive in language and other cognitive processes, as recent studies demonstrate (Loewenstein and Gentner 2005, Perrott et al. to appear).

[7]There is also accumulating linguistic evidence that 'words' (the irregular lexicon) and 'rules' (phrase-structure rules) are not a dichotomy but rather, a continuum. Cf. Culicover's 'syntactic nuts' (Culicover 1999, Culicover and Jackendoff 2005).

Moreover, the process of comparison inherent in the recognition of analogies has been shown to facilitate the abstraction of rules in young children (Gentner and Medina 1998).

On the other hand, several neural-network computational approaches have been able to successfully model exemplar-based learning and subsequent productivity (McClelland et al. 1986, Regier et al. 2001, Gasser and Colunga 2002, Plaut and Gonnerman 2000, Chang 2004, Chang and Gurevich 2004). Finally, in the domain of morphology, Baayen and his colleagues have demonstrated the importance of entrenchment and numbers of morphologically related words ('derivational families') for morphological productivity (Baayen and Moscoso del Prado Martín 2005, Hay and Baayen 2005, Hay 2001, Baayen et al. 2003, Moscoso del Prado Martín et al. 2005).

The combined weight of the psycholinguistic and computational studies just mentioned provides strong evidence in favor of distributed analogy as a plausible mechanism for productivity, both in syntax and in morphology. So, while we can sketch symbolically how an abstract pattern may extend to novel items, we should keep in mind that productivity really does seem to be example-based rather than abstractly symbolic.

Another important point to be made is that analogy is inherently *probabilistic* in nature. Both the morphological studies and the connectionist models just cited provide evidence that analogical extension works on the principle of best fit. There is not always an obvious choice for analogical host, and sometimes there may be multiple possibilities. When there is a choice of pattern for analogical extension, the decisions appear to be made based on relative frequency and what Bybee (1985, 2001, 2001) has called relative 'entrenchment' of linguistic patterns. Thus, analogical models are in principle compatible with stochastic Optimality Theory models in phonology and syntax (Prince and Smolensky 1993, Legendre et al. 2001), though without assuming universal constraints.

2.5.2 Well-formedness and Negative Evidence

Another important question for a model of grammar is how it assures well-formedness of words and phrases that are produced. Generative grammars have an explicit goal of producing all and only grammatical words and sentences. A grammar consisting of a set of example-based licensing constructions, on the other hand, may have a tendency to overgenerate if constructions are freely combined with each other. How, then, is the issue resolved in construction-based approaches?

Part of the well-formedness in inflectional morphology can come from morphotactic and templatic constraints on what constitutes a whole word form. Another is constraint comes from the way in which analogical hosts are selected: even if one starts at the abstract top of the constructional hierarchy, with a set of semantic properties to express, one must travel down it to find one (or more) fully specified, compatible constructions from which to extend to the novel item.

Also, higher-level constructions (e.g. agreement or case) may require otherwise well-formed words to participate in them. A word (construct) that instantiates multiple constructions may instantiate a higher-level construction, which in turn requires compatibility with several other constructions specifying the root, stem, etc. So unification is important, but is not by itself sufficient to guarantee well-formedness. It is not true that any two constructions can just be stuck together and unified if they are compatible; most of the time, a higher-level construction licensing the two participants must exist.

In general studies in language acquisition provide evidence for exemplar-based methods for constraining linguistic productions. In the production of syntactic frames, Pinker (1989) has suggested several possible mechanisms. First, children may be more likely to extend a verb beyond its habitual use the more frequent the habitual constructions are (*entrenchment*); this has also been proposed by (Clark 1987, Goldberg 1995, Bates and MacWhinney 1989). Subsequent studies provided empirical evidence for entrenchment (Brooks et al. 1999, as described in Tomasello 2006), particularly in the earlier stages of acquisition. Second, children may be sensitive to semantic classes of verbs and fail to extend their use to situations that call for different meaning; this too has been empirically confirmed by (Brooks and Tomasello 1999) in somewhat older children. Third, children may be *preempted* from extending a construction if they have heard an alternative construction used in the same communicative function. The experiments in (Brooks et al. 1999, Brooks and Tomasello 1999) provide some evidence of preemption as well. Given the emerging body of evidence for exemplar- and frequency- based acquisition and processing of languages, it seems that exemplar-based models in linguistic theory are well justified.

2.6 Case Studies

The rest of this dissertation presents several case studies that illustrate the mechanisms of Constructional Morphology and elaborate them where needed. The case studies

are as follows.

1. Georgian verbal morphology (Chapter 3).

   This case study deals mostly with inflectional morphology and its relation to syntax. The main challenges handled by the analysis are subject and object agreement, heterogeneous case-marking on verbal arguments, and the distribution of inflectional elements on verb forms. The chapter suggests that the tense/aspect/mood series in Georgian is the central organizing unit that mediates between the divergent patterns in syntax and morphology, and is best analyzed as a construction. Crucially, the analysis involves a hierarchy of constructions at various degrees of abstractness.

2. Georgian version (Chapter 4).

   This case study examines the phenomenon of version, which was briefly introduced in Chapter 1. The chapter focuses on syntactic and semantic properties of pre-radical markers (i.e. version 'proper'). The different uses of version markers are analyzed as a family of related constructions that involve valence alternations. The chapter extends the theoretical framework by adding a mechanism to handle valence alternations. A representation for relations between constructions that do not involve inheritance is introduced, relying on paradigmatic contrasts.

3. Morphosyntactic uses of pre-radical vowels (Chapter 5).

   This case study is an example of a morphological distribution that does not have a semantic correlate. The function of pre-radical vowels is determined by paradigmatic contrasts. The theoretical mechanism involves representation of contrasting constructions.

4. Computational Modeling (Chapter 6).

   This chapter describes a computational model for parsing and generating a subset of inflected Georgian verb forms. It is a case study in applying the results of a constructional analysis to computation, which typically relies on compositional rules.

At each step, efforts will be made to reconcile the theoretical analyses with existing psycholinguistic and acquisition evidence.

Chapter 3

Constructions in Georgian Morphosyntax

3.1 Introduction

Georgian exhibits complex syntactic and morphological patterns that may at first glance seem irregular and unpredictable. Grammatical relations such as subject and object appear to be reflected differently and inconsistently by case-marking and verbal agreement. Case-marking and agreement systems are themselves dependent on the conjugation class and tense of a particular verb, and as a result the relationship between morphological forms and grammatical relations is obscured, leading some scholars to question the usefulness of notions subject and object for Georgian. In addition, verbs have complex morphology in which the same formal units (morphs) recur in different arrangements in different contexts, so that it is virtually impossible to assign individual meaning to them. Yet upon closer examination, Georgian morphosyntax is highly regular, and some of its key features have been stable since the earliest attestations of Old Georgian in the 5th century AD (Harris 1985).

The purpose of this chapter is to clarify the status of the recurrent patterns in Georgian and their relation to grammatical relations and agreement. The analysis consists of a lexical and paradigm-based specification of the conjugation classes and a constructional representation of the Tense-Aspect-Mood (TAM) system, including nominal case marking and agreement. I will argue throughout that there is a natural unit of organization in

Georgian, namely the TAM series, which mediates between case-marking, agreement, and conjugation class patterns and is best thought of as a construction. I will begin with a brief overview of the main issues in analyzing the Georgian verbal system, followed by a more detailed examination of the morphosyntactic patterns recurrent in the language.

First, a short note on terminology is in order. Throughout this chapter the the following chapters on version, the relationship between thematic roles and syntactic arguments will play an important role. Theories differ in the names and functions they assign to different thematic roles. Since the main interest here will be the *linking* between thematic roles and syntactic arguments, the particular names chosen are not crucial. I will use the terms *Agent* and *Patient* in the most pre-theoretical way possible, simply to refer to scene participants that differ in their degree of volitionality, intention, affectedness, etc. Dowty (1991) defines the concept of *proto-roles*, the occurrence of which entails clusters of such properties. Ackerman and Moore (2001) refine the distinctions further, but I will not address the specific differences in these two approaches here. Terms such as Agent and Patient can also be compared to *logical subject* and *logical object* in Lexical Functional Grammar and, specifically, lexical mapping theory (Bresnan 2001:307), or to *initial subject* and *initial object* in Relational Grammar (Perlmutter 1983). For the purposes of the present dissertation, the *paradigmatic* comparison of linking constructions is more important than the individual properties of the thematic roles.[1] In addition, the term *Goal* is used to represent the third type of thematic role (not identified as a proto-role by Dowty), to indicate a third participant most often associated with the indirect object.

The basic issue in the Georgian case-marking system is as follows. The agent (logical subject in LFG terms) in different constructions can be marked by the Nominative case, as in (53a), the Narrative case (53b)[2], or the Dative case (53c).

(53) a. *k'ac-i* *dzaγl-s* *xat'avs*
man.NOM dog.DAT paint.PRES.3SGSUBJ.3OBJ
'The man paints / is painting the dog.' (Present)

[1]See Ackerman and Moore (2001) for a thorough discussion of syntagmatic vs. paradigmatic selection of arguments.

[2]Narrative is the traditional name for this case. It has also been called Ergative, implying a corresponding ergative analysis of the Georgian case system (Hewitt 1987a, 1987b, 1995a). However, as Harris (1981, 1985, 1989, 1990) has argued, there is significant evidence that the synchronic case-marking system follows an active-stative rather than ergative-absolutive pattern. While I am inclined towards the latter view, I wish to remain neutral about this debate here and use the traditional case name.

b. *k'ac-ma dzaγl-i daxat'a*
man.NAR dog.NOM paint.AOR.3SGSUBJ.3OBJ
'The man painted the dog.' (Aorist)

c. *k'ac-s dzaγl-i turme dauxat'avs*
man.DAT dog.NOM apparently paint..PERF.3SGSUBJ.3OBJ
'The man has apparently painted the dog.' (Perfect)

The patient (logical direct object) in different constructions can be marked by the Dative case, as in (53a) above, or the Nominative case, as in (53b) and (53c) above. The goal (logical indirect object) in different constructions can be marked by the Dative case as a core argument (54a) or appear in a postposition phrase (54b). The last pattern is known as "inversion" because of the apparently reverse marking of subjects and indirect objects.

(54) a. *k'aci c'igns p'avles chuknis*
man.NOM book.DAT Paul.DAT give.PRES.1SGSUBJ.3OBJ
'The man is giving a book to Paul'. (Present)

b. *k'acs c'igni p'avlesa-tvis uchukebia*
man.DAT book.NOM Paul.GEN-for give.PERF.1SGSUBJ.3OBJ
'The man has given a book to Paul'. (Perfect)

There is no direct mapping between case and grammatical relations; rather, the mapping is determined by the conjugation class and TAM of the verb, as will be illustrated below.

The verbal agreement system presents a somewhat separate set of issues. A verb can cross-reference up to two arguments at a time using pre- and post-stem agreement markers. The distribution of the agreement morphs interacts with conjugation class and verb tense, and the resulting mapping between agreement morphology and participant roles is often apparently variable. So, in (55) below, the marker *g-* indicates a 2Sg *agent*, whereas in (56) the same marker *g-* indicates a 2Sg *patient*.

(55) *me shen g-xat'av*
I you 2SG-paint.PRES
'I am painting you.' (Present)

(56) *shen me da-g-i-xat'ivar*
you I PV-2SG-paint.PERF
'You have (apparently) painted me.' (Perfect)

Despite the different mapping to thematic roles, these agreement morphs have consistent associations with person/number properties (and possibly also with grammatical functions, on the traditional analysis given in Tschenkéli (1958)) and are thus recurrent patterns. The analysis proposed below brings many of the different factors affecting the mappings between morphological form and grammatical relations into a single framework. I will argue that a representation of conjugation classes as lexical and a constructional description of the TAM and valence patterns resolve the apparent inconsistencies of the Georgian morphosyntax.

I will start with a brief description of the Georgian TAM system, followed by a description of the four conjugation classes and case-marking patterns for different conjugation classes in different tenses. A discussion of the verbal agreement system and related issues will follow. In all cases, I will argue that there is a natural organizational unit around which the recurrent case-marking and agreement patterns cluster, namely the *series*.

3.2 Morphosyntactic Patterns

3.2.1 Screeves and series

A Georgian verb can appear in up to ten semantically distinct form sets based on TAM properties[3]. These sets, which roughly correspond to traditional inflectional paradigms, are called **screeves**, from Georgian *mc'k'rivi* 'row'. Screeves have traditionally been grouped into three series based on commonalities in morphological formation of the forms themselves and the associated case-marking and agreement patterns. The distribution of screeves in series is described in Table 3.1.

SERIES I (IMPERFECT)		SERIES II	SERIES III
PRESENT SUBSERIES	FUTURE SUBSERIES	(AORIST)	(PERFECT)
Present	Future	Aorist	Perfect
Imperfect	Conditional		Pluperfect
Present subjunctive	Future subjunctive	Optative	(Perfect subj.)

Table 3.1: Series and Screeves ("Present", "Future", etc. are screeves)

[3]An eleventh form set, the Perfect Subjunctive, is no longer used in modern Georgian except in poetry or archaic expressions (cf. Hewitt 1995b, Kurtsikidze forthcoming)

All screeves in a given series share agreement and case-marking patterns. In addition, there are systematic morphological correspondences between the screeves in the same column. So, the Future subseries in most cases differs from the Present subseries by having a lexically specified preverb (Future screeve corresponds to Present screeve, Conditional corresponds to the Imperfect, Future Subjunctive corresponds to the Present Subjunctive). Variations in stem selection of Future forms predict variations in the Aorist. However, there are also correspondences that are not aligned by column. So, some types of variation in Optative screeve formation predicts variation in the Pluperfect.

This distribution already makes apparent certain TAM properties common to series. Series I contains the traditionally imperfective, durative screeves, and their perfective counterparts[4]. Series II contains the past perfective screeves, and series III contains the present and past perfect tenses, commonly used as evidentials. It will be illustrated below that the series also mediates between different morphosyntactic patterns. In addition, series and screeves themselves have a complex of formal, syntactic and semantic properties that a construction-based account is ideally suited to handle.

3.2.2 Conjugation Classes, Valence, and Case-Marking

Traditionally, Georgian verbs have been grouped into four conjugation classes. These classes are best defined as formal groupings based on patterns of word-formation and agreement. Many previous attempts have been made to describe the conjugation classes as semantically or syntactically based (Hewitt 1995b, Cherchi 1997) but inevitably each class contains some exceptional members that do not match the semantic criteria. This state of affairs is by no means unusual cross-linguistically: inflectional classes are defined most reliably in terms of form, but are often cued by secondary syntactic and semantic factors. For example, the Slavic declension classes discussed in Section 2.4.2 are defined on the basis of case endings, but reinforced by fairly consistent gender associations. The association between conjugations and valence in Georgian is just another example of this. The criteria used here are largely the same as those used by Harris (1981) and include the formation of the Future and Aorist series and the marking of 3rd person subjects in various screeves. Table 3.2 summarizes the distinguishing characteristics of the conjugation classes.

Case-marking of nominal arguments varies across conjugation classes. However,

[4]The most comprehensive analysis of Georgian aspect is Holisky 1981a, Holisky 1981b.

	Conjugation 1	Conjugation 2	Conjugation 3	Conjugation 4
Syntactic class	Transitive	Unaccusative	Unergative	Inverse
Typical valence	Transitive, active subjects	Intransitive, often passive derived from C1	Intransitive, active subjects	Transitive, experiencer subj.
Future, Aorist	Preverb	Preverb or *e-*	Circumfix *i–(eb)*	*e-*
Future 3Sg Subj	*-s*	*-a*	*-s*	*-a*
Future scr. vowel	-	*-i*	-	-
Aorist 3Pl Subj	*-es*	*-nen*	*-es*	Same as 3SgSubj
Examples	*gazrdis* 'he will raise X'	*gatq'deba* 'it will break'	*icek'vebs* 'he will dance'	*eq'vareba* 'he will love X'

Table 3.2: Conjugation classes

there are some recurrent patterns even in these seemingly inconsistent patterns. The three possible case-marking patterns were demonstrated in examples (53a-53c) above, using the Conjugation 1 verb *xat'va* 'to draw'. Tables 3.3 - 3.6 show the case marking characteristics for different series of each conjugation class, using thematic roles rather than syntactic arguments for reasons that will become clear when we discuss inversion.

	Agent	Patient	Goal
series I	NOM	DAT	DAT
series II	NAR	NOM	DAT
series III	DAT	NOM	oblique

Table 3.3: Conjugation 1 case marking

Generally, all screeves in a given series for a given conjugation class will follow the same case-marking pattern. There are essentially only three patterns by which the participants can be marked. If the agent is marked by the Nominative, the patient can only be marked by the Dative; if the agent is marked by the Narrative, the patient can only be marked by the Nominative; and if the agent is marked by the Dative, the patient will always be marked by the Nominative. These three patterns have often been represented as in 3.7 (based on marking of agent, following Harris (1981)).

The general issues with the case-marking system are as follows. First, there is no

	Agent	Goal
series I	NOM	DAT
series II	NOM	DAT
series III	NOM	DAT

Table 3.4: Conjugation 2 case marking

	Agent	Goal
series I	NOM	DAT
series II	NAR	DAT
series III	DAT	oblique

Table 3.5: Conjugation 3 case marking

default way to mark a single argument of a verb. The single argument could be Nominative (as in Conj II), Narrative (as in Conj III, series II), or Dative (as in Conj III, series III).

Second, cases do not have consistent meanings throughout the different series and conjugation classes. The Nominative can mark the subject (in series I of Conjugations 1 (active transitive) and 3 (medioactive)) or the direct object (in series II and III of Conjugation 1 (active transitive)). The Dative can mark the logical subject (in series III of Conjugation 1), the direct object (in series I of Conjugation I), or the indirect object (in series I and II of Conjugations 1, 2 (passive), and 3).

Many syntactic traditions rely on semantic argument rankings in order to assign grammatical relations and case (for example, the linking between thematic roles and arguments in Lexical-Functional Grammar, where the highest-ranking thematic role becomes the subject). An example thematic hierarchy is given in 57.

(57) Thematic Hierarchy (from (Bresnan 2001:307)):
agent $\succ$ *beneficiary* $\succ$ *experiencer/goal* $\succ$ *instrument* $\succ$ *patient/theme* $\succ$ *locative*

The simplest form of a hierarchical approach states that the subject is linked to the highest-ranking thematic role present in the clause, with morphological properties like case-assignment and agreement following in step. Clearly, this simple role assignment would not work for Georgian, both because the highest-ranking thematic argument is not always the subject (Pattern C), and because morphological case-marking is different in patterns

	Agent	Patient	Goal
series I	DAT	NOM	oblique
series II	DAT	NOM	oblique
series III	DAT	NOM	oblique

Table 3.6: Conjugation 4 case marking

Table 3.7: General case-marking patterns

	Agent	(Patient)	(Goal)
A	NAR	NOM	DAT
B	NOM	DAT	DAT
C	DAT	NOM	postposition

A and B. Of course, most theories of syntactic argument structure are more sophisticated and allow for alternative linking rules that account for passives, applicatives, and other alternations (Bresnan 2001, Ackerman and Moore 2001, and many others). More specific linking proposals for Georgian are discussed in Section 3.2.6.

The greatest problem presented by Georgian is specifying when a particular linking pattern is to be used. As the examples above demonstrate, the choice of pattern is determined by conjugation class and screeve/series of a verb. In this respect, the Georgian linking patterns differ from alternations such as the English passive, which is assumed to apply via lexical rules to any transitive verb stem (see Sag et al. 2003, ch.10 for an example HPSG account). Rather, it depends on a combination of a lexical property (conjugation class) and syntactic/semantic construction (series/screeve).

The inconsistencies in case marking and agreement have led some to question the usefulness of grammatical notions like 'subject' for Georgian. Tuite (1994) applies several morphological and syntactic tests for subjecthood, such as case- and agreement- assignment and anaphora control, in standard Georgian and several of its dialects and concludes that standard Georgian does, indeed, have the notion of 'subject,' though not very prominently; and some of the regional dialects do not make salient use of this notion at all. Aronson (1984, 1994) examines cases of (surface) syntactic homophony and lability frequent with some verb classes and suggests that syntagmatic information from sentence forms is often not sufficient to interpret the sentence and determine the grammatical arguments in it.

Noting, however, that speakers have no trouble understanding each other, he suggests that speakers must use paradigmatic information on how the sentences *may* have been encoded, as well as contextual information, to interpret sentences. More generally, Aronson suggests that the universalists notions of subject and object, while applicable to Georgian to some extent, do not serve well to explain how speaker interaction occurs, and should be looked at from a different perspective. It seems that a constructional, exemplar-based approach, could provide firmer foundations for how sentences are produced and comprehended by speakers without the advantage of clearly defined grammatical relations.

There are, however, useful regularities in the system. In all cases, the series for a given conjugation class determines the case-marking pattern, and there are altogether only three such patterns. The following two sections examine verbal agreement corresponding to the case-marking patterns and suggest that agreement, too, relies on linking patterns determined at the level of the series.

3.2.3 Verb Agreement: Direct

A Georgian verb can agree with its subject, object, indirect object, and occasionally other arguments, although some constraints apply on the particular combination of arguments and their properties. Verbal agreement is expressed by affixes that appear before or after the verb stem. Many grammatical descriptions divide verbal agreement markers into different sets based on which syntactic argument they refer to. Some agreement markers depend on the conjugation class of a verb, as mentioned above. However, a general set of agreement markers can also be deduced that does not depend on conjugation class. Following Harris (1981) and other scholars, I will identify these sets by their salient affix, as in Tables 3.8, 3.9, and 3.10.

	SINGULAR	PLURAL
1st Pers	*v-*	*v-...-t*
2nd Pers	*ø-*	*ø-...-t*
3rd Pers	*ø-...-s/-a/-o*	*ø-...-an/-en/-es/-nen*

Table 3.8: *v*-set (subjects of non-inverted verbs)

Subject and direct object agreement are the most common and productive types.

The *v*-set marks subjects of non-inverted verbs (that is, series I and II of Conjugations 1 and 3, and all series of Conjugation 2). The 3rd person singular and plural affixes vary depending on the conjugation class and screeve and will be described in more detail in the next section. The *m*-set marks direct objects of non-inverted verbs, and the *h*-set marks indirect objects of non-inverted verbs. The *h*-set and *m*-sets are identical save for the *h-/s-* prefix on the 3P forms of indirect objects. In some sense, the markers *m-*, *g-*, *gv-*, and *-t* mark an object regardless of whether it is direct or indirect, whereas the *h-/s-* prefixes (which are phonologically conditioned variants) specifically index indirect objects.

	SINGULAR	PLURAL
1st Pers	m-	gv-
2nd Pers	g-	g-...-t
3rd Pers	ø-	ø-...-t

Table 3.9: *m*-set (direct objects of non-inverted verbs)

	Sing	Plur
1P	m-	gv-
2P	g-	g-...-t
3P	h-/s-/ø-	h-/s-/ø-...-t

Table 3.10: *h*-set (indirect objects of non-inverted verbs)

The two sets of agreement markers combine in a somewhat complex manner. A verb form can only have one agreement prefix, resulting in a kind of slot competition when a verb form makes two markers semantically possible, e.g. *v-* and *g-* in a 1st person subject – 2nd person object form. This competition is always resolved in favor of the *m-* or *h-* set marker; in other words, the object markers 'win'. In the suffixal slot, only one *-t* is allowed, and this marker shows interesting slot interactions with some of the preceding screeve markers (see Section 3.3.4 for a discussion of these interactions). The primacy of object over subject is consistent with the historical pattern stemming from old Georgian (Chikobava 1984). Since the pronominal agreement markers are clearly identifiable from full forms, several technical morphological accounts have attempted to assign meanings, or realizational rule correspondences, to these elements (Anderson 1992, Halle and Marantz 1994,

Stump 2001)[5]. However, as section 3.3 aims to demonstrate, even a successful assignment of meaning and rule ordering to these elements does not resolve the larger issues in the morphotactics of Georgian verbs. Thus, agreement markers will not be discussed here in great detail.

The present-tense paradigm of the verb *xat'va* 'to paint' (Conj 1) in Table 3.11 illustrates how the *v*-set and *m*-sets interact. Reflexivity in Georgian is expressed using the 3rd person reflexive pronoun *tavi* (meaning 'head'; cf. Amiridze 2006), and so forms with identical subject and object persons do not exist, as indicated by the dashes in the table.

AGENT (NOMINATIVE)	PATIENT (DATIVE)				
	1SG	1PL	2SG	2PL	3
1SG	—	—	***g***-*xat'av*	***g***-*xat'av*-***t***	***v***-*xat'av*
1PL	—	—	***g***-*xat'av*-***t***	***g***-*xat'av*-***t***	***v***-*xat'av*-***t***
2SG	***m***-*xat'av*	***gv***-*xat'av*	—	—	*xat'av*
2PL	***m***-*xat'av*-***t***	***gv***-*xat'av*-***t***	—	—	*xat'av*-***t***
3SG	***m***-*xat'av*-***s***	***gv***-*xat'av*-***s***	***g***-*xat'av*-***s***	***g***-*xat'av*-***t***	*xat'av*-***s***
3PL	***m***-*xat'av*-***en***	***gv***-*xat'av*-***en***	***g***-*xat'av*-***en***	***g***-*xat'av*-***en***	*xat'av*-***en***

Table 3.11: *xat'va* 'to paint' in the Present

The precise way in which the two sets of affixes combine in verb forms is a matter of considerable interest and complexity and has often been analyzed in terms of slot competition between rules realizing subject and object properties (Anderson 1992, Stump 2001). However, the existing slot analyses fail to take into account the full range of relevant verb forms and do not offer a fully general solution. As will be shown below, morpheme-based and even realizational, rule-ordering analyses of Georgian verbal morphology are not ideal for capturing the recurrent patterns in the system.

3.2.4 Verb Agreement: Inversion

In certain conjugation classes and screeves, the cross-reference of verbal arguments appears to be reversed. This reverse agreement marking, called *inversion*, happens in series III (i.e. the perfect tenses) of Conjugation 1 ("transitive") and 3 ("unergative") verbs, and

[5]See (Harris 2004) for a historically-based argument against all three accounts.

in all series of Conjugation 4 ("indirect") verbs. Compare the Present paradigm for *xat'va* 'to paint' in 3.11 with the Perfect paradigm of the same verb in 3.12. The formation of the perfect incorporates a form of the copula 'to be' agreeing in person with the nominative argument (*var*, *xar* are forms of the copula). For convenience, both tables are organized based on case-marking: the Nominative argument is in rows ("agent" and "surface" subject in Present; "patient" / "surface" subject, in Perfect) and the Dative argument is in columns ("patient" and "surface" object in Present, "agent" / "surface" object in Perfect).

PATIENT (NOM)	AGENT (DATIVE)			
	1SG	1PL	2SG	2PL
1SG	—	—	*da-**g**-**i**-xat'i-**var***	*da-**g**-**i**-xat'i-**vart***
1PL	—	—	*da-**g**-**i**-xat'i-**vart***	*da-**g**-**i**-xat'i-**vart***
2SG	*da-**m**-**i**-xat'i-**xar***	*da-**gv**-**i**-xat'i-**xar***	—	—
2PL	*da-**m**-**i**-xat'i-**xart***	*da-**gv**-**i**-xat'i-**xart***	—	—
3	*da-**m**-**i**-xat'av-**s***	*da-**gv**-**i**-xat'av-**s***	*da-**g**-**i**-xat'av-**s***	*da-**g**-**i**-xat'av-**t***

PATIENT (NOM)	AGENT (DATIVE)	
	3SG	3PL
1SG	da-**u**-xat'i-**var**	
1PL	da-**u**-xat'i-**vart**	
2SG	da-**u**-xat'i-**xar**	
2PL	da-**u**-xat'i-**xart**	
3	da-**u**-xat'av-s	da-**u**-xat'av-**t**

Table 3.12: *xat'va* 'to paint' in the Perfect

An interesting issue arises when these general agreement markers are mapped onto grammatical relations and compared to case-marking patterns discussed above. Table 3.13 illustrates the mapping between thematic roles and agreement marker sets for the different conjugations and series, displayed alongside the case-marking patterns.

It seems that the mapping between agreement marker sets and thematic roles changes only in cases of inversion. On the other hand, the mapping between cases and grammatical relations varies with conjugation class and series. There is no consistent association between agreement marker sets and case. Agreement morphs have consistent associations with person/number properties and a variable association with thematic roles. However, agreement patterns are consistent throughout the series for each conjugation class.

	Agent	(Patient)	(Beneficiary)
A (narrative)	NAR, v-set	NOM, m-set	DAT, h-set
B (direct)	NOM, v-set	DAT, m-set	DAT, h-set
C (inverse)	DAT, h-set	NOM, v-set	oblique, –

Table 3.13: Case-marking and agreement patterns

Table 3.14 demonstrates the distribution of the three patterns (A, B, and C) throughout conjugation classes and series. The next section briefly examines some phenomena that are independent of linking but sensitive to thematic roles instead.

SERIES	CONJ 1	CONJ 2	CONJ 3	CONJ 4
Series I	B	B	B	C
Series II	A	B	A	C
Series III	C	B	C	C

Table 3.14: Case and Agreement Patterns by Conjugation Class and Series

3.2.5 Phenomena Sensitive to Thematic Roles

Several syntactic phenomena in Georgian remain sensitive to thematic roles (or "deep" grammatical relations, in RG terminology). These include number agreement, reflexivization, and several others. In general, only animate arguments trigger plural agreement in Georgian. In addition, there are important differences in plural agreement between non-inverted and inverted screeves. Compare again the Present and Perfect screeve paradigms of 'paint' with 3rd person surface subjects (nominative arguments) (Table 3.11 vs. 3.12).

In the Present paradigm, the 3Pl Nominative argument (agent) triggers agreement on the verb with the suffix *-en*. There is no plural agreement for a 3rd person Dative argument (patient). In the Perfect paradigm, by contrast, there is separate plural agreement for the Dative argument (agent) but not for the Nominative argument (patient).

The distribution of most agreement morphs (*m-*, *g-*, *gv-*) appears to depend on surface grammatical relations, signified by case. However, the absence vs. presence of plural agreement is sensitive to the thematic roles; only semantic agents can trigger plural

agreement in 3rd person. This fact points to a need for access to semantic as well as syntactic information when selected verbal inflections are formed.

Reflexivization is another syntactic phenomenon that is sensitive to thematic roles rather than surface grammatical relations. Reflexivization is expressed by a noun phrase headed by the reflexive pronoun *tavi* 'self, head' and is triggered by the Nominative argument (semantic agent) in series I of transitive verbs (58a) and by the Dative argument (also semantic agent) in series III (58b).

(58) Inversion and Reflexivization (Harris 1981:125)

a. *gela* *irc'munebs* *tavis* *tavs.*
Gela.NOM convince.FUT self.GEN self.DAT
'Gela is convincing himself.'

b. *gelas* *turme* *daurc'munebia* *tavisi* *tavi.*
Gela.DAT apparently convince.PERF self.GEN SELF.NOM
'Gela apparently (has) convinced himself.'

Reflexivization, along with unemphatic pronoun drop and a set of other syntactic diagnostics, has been used to argue that the Dative arguments in Inversion are still "deep" subjects in some sense (Harris 1981, Tuite 1994), but in the terminology used here, this is equivalent to saying that the Dative argument is still the semantic Agent, and reflexivization is triggered by the thematic role rather than the surface grammatical relation. The following section presents a linking analysis of the case and agreement mismatches.

3.2.6 Case-Marking and Agreement as Linking Problems

Harris (1981) argues that Pattern C above is a subtype of Pattern B with a different linking of case-marking and morphology to logical syntactic relations. She claims that the Dative argument in Pattern C starts out as an initial subject but is realized as a surface indirect object, thus receiving Dative case and *h*-set agreement marking. Working within the derivational framework of Relational Grammar, she argues for a two-level syntactic representation. Anderson (1984) criticizes Harris' derivational syntactic approach and instead suggests that the derivations happen in the morphology. However, he retains the idea that inverted and non-inverted clauses 'start out' with the same syntactic structure (Harris' initial structure) and, through some sort of derivation, end up exhibiting the inverse case-marking and agreement effects.

A derivational approach seems committed to the view that once the derivation has happened, surface forms should no longer be able to access properties of the initial, or deep-syntactic, forms. However, as the previous section demonstrated, Georgian remains sensitive to some properties of 'initial' subjects and objects for purposes of number agreement and some others. On the other hand, the insight that subjects in inversion have some properties of non-inverted indirect objects, and inverted objects are in some sense similar to non-inverted subjects, is clearly valid, and is salient in traditional grammatical descriptions of Georgian such as (Tschenkéli 1958) and others.

Contemporary mapping theories such as Lexical Mapping Theory (LMT; Bresnan and Zaenen 1990, Bresnan 2001) establish a direct, single-level linking between semantic roles and syntactic arguments, providing a means of capturing the traditional intuition without assuming that semantic roles have temporal priority over syntactic arguments. In a single-level representation, there is no *a priori* reason why both types of information could not be used by agreement and case-marking rules. A linking analysis, such as the one in (Blevins to appear), proposes two different linking patterns, one with a 'default' mapping of agent to subject and patient to object, and one with an 'inverted' mapping of agent to (indirect) object and patient to subject. The linking patterns are associated with verbal lexical entries and stipulate the mapping between the semantic agent and a grammatical function (subject or object), as in (59). The assignment of the rest of the thematic roles takes place via a set of default rules based on the thematic hierarchy in (57) and a feature geometry in which all grammatical functions are marked as [±*r(estricted)*] and [±*o(bjective)*] (Bresnan 2001:308-9). The precise mechanism of default linking is not crucial to the analysis of inversion and will not be explained in detail here.

(59) Direct and inverted linking for the verb *xat'va* 'paint' (based on Blevins to appear; GF = Grammatical Function)

a.

	xat'va	<*ag*	*th*>
intrinsic:			[-*r*]
GF:		SUBJ	OBJ

b.

	xat'va	<*ag*	*th*>
intrinsic:			[-*r*]
inverse:		|	
GF:		IND-OBJ	SUBJ

Linking approaches may vary in how this assignment occurs: just one of the links may be stipulated, as in the example above, or all of the links could be made explicit, as in (Goldberg 1995). The default, thematic hierarchy-based approach has more cross-linguistic explanatory power, as it may account for the common patterns of argument selection cross-linguistically. However, for the purposes of keeping constructional representations more readable, I will adopt something closer to Goldberg's notation, with the full understanding that the proposed linking rules could be expressed within Lexical Mapping Theory as well.

To reiterate, the Georgian Patterns B (direct) and C (inverse) are basically similar in how syntax maps to case, but different in how semantics maps to syntax. In Pattern B, the mapping is from Agent to subject and from Patient to direct object, whereas in Pattern C, the mapping is from Agent to indirect object and from Patient to subject. One pattern is not derived from the other in any sequential way; rather, the difference between them is in the mappings between semantic roles and syntactic arguments. Thus, as Harris argues, Patterns B and C can be viewed as two instances of the same general pattern where the subject is marked by the Nominative and *v*-set agreement, and the indirect object is marked by the Dative and *h*-set agreement. These patterns can be represented as subcases of a general Pattern 1 in Table 3.15.

Default Mapping (Pattern B):	**Inverse Mapping** (Pattern C):
Agent → *subj*; *Patient* → *DO*; *Goal* → *IO*	*Agent* → *subj*; *Patient* → *IO*; *Goal* → *oblique*
Pattern 1 case-marking and Agreement:	
Subj → *NOM*, *v* − *set*; *DO* → *DAT*, *m* − *set*; *IO* → *DAT*, *h* − *set*; *Non* − *core* → *oblique*	

Table 3.15: Pattern 1: Nominative/Accusative

Pattern 1 displays the standard Nominative/Accusative treatment of subject and object. The sensitivity of 3Pl agreement to semantic rather than syntactic roles suggests that morphs do not simply 'realize' syntactic properties. Rather, the morphology is sensitive to syntactic *as well* as semantic information, and there is no sense in which the syntactic information is uniformly derived from semantics.

Pattern A, with the narrative-marked subjects, is the exceptional pattern, then. Harris (1985) has shown that this is historically the oldest pattern in Kartvelian, going

back to the time when Georgian was fully ergative. She demonstrates that this pattern has been reanalyzed as active rather than ergative, such that active agents of verbs in series II are marked by the ergative/narrative case (cf. King 1994). This claim is supported by the fact that only subjects of Conjugation 1 (transitive) and Conjugation 3 (mostly active intransitive / unergative) are marked by NAR, while the inactive subjects of Conjugation 2 (intransitive / passive / unaccusative) verbs are marked by NOM case. So Pattern A applies only to Conjugation 1 and 3 verbs, in series II. This pattern can be termed Active Pattern 2 and represented in Table 3.16.

Default Mapping (Pattern A):			
$Agent$	$Patient$	$Goal$	
$\downarrow$	$\downarrow$	$\downarrow$	
$subj$	DO	IO	
Pattern 2 case-marking and Agreement:			
$Subj$	DO	IO	$Non-core$
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
$NAR, v-set$	$NOM, m-set$	$DAT, h-set$	$oblique$

Table 3.16: Pattern 2: Active Pattern 2 case-marking and Agreement

Historically, the default NOM/ACC pattern (Harris' pattern B, or the direct-linking subcase of Pattern 1 above) arose from the Ergative pattern through reanalysis of an antipassive construction. This is unsurprising particularly because the former Absolutive case became reanalyzed as Nominative, and both cases are often unmarked cross-linguistically. Synchronically, however, both the active pattern and the inverse patterns seem to be deviations from the Default Nominative/Accusative pattern. The Nom/Acc pattern is the basic and most frequent pattern in Modern Georgian. The inverse pattern differs from it in how semantic participants are linked to syntactic arguments, and the active pattern differs in formal case assignment for the syntactic arguments but retains the same linking to semantic participants and agreement. The choice of pattern is, again, determined by the series and conjugation class of a verb. Table 3.14 is repeated in Table 3.17 with Patterns 1 and 2 instead of the more detailed distinctions of Patterns A, B, and C.

From an analyst's point of view, one of the problems of the Georgian system is the variable use of cases so that their meanings are difficult to determine. On the account suggested above, the Narrative case can be characterized as marking "active subjects in series II", and the Dative consistently marks "surface Direct or Indirect Objects". The

SERIES	CONJ 1	CONJ 2	CONJ 3	CONJ 4
Series I	1	1	1	1 (inverse)
Series II	2	1	2	1 (inverse)
Series III	1 (inverse)	1	1 (inverse)	1 (inverse)

Table 3.17: Case and Agreement Patterns by Conjugation Class and Series

Nominative remains a bit of a mystery as it marks subjects in Pattern 1 and direct objects in Pattern 2. It can be seen as a kind of default case that marks the argument not already marked by one of the unambiguous cases. So, in Pattern 2, the Nominative is used to mark the direct object because the subject is already unambiguously marked by Narrative, and in Pattern 1 the Nominative is used to mark the subject because the pattern does not specify any other marking for the subject. Alternatively, one could say that the Narrative marks active subjects in Series II, the Dative marks indirect objects directly governed by a verb, and Nominative is associated with the highest (least marked in LFG terms) grammatical function that is not associated with a case by the first two principles. This 'default' interpretation of the Nominative goes along with the fact that it is the least morphotactically marked case in Georgian: it requires the addition of a single vowel *-i* to consonant-stem nouns, and no modification at all for vowel-stem nouns and is the citation form for nouns.

In addition to the different mappings between thematic roles and syntactic arguments, inversion exhibits certain constructional properties that are motivated, but not necessarily predictable from, the different mapping. For instance, the difference in plural agreement is motivated by semantic properties of agents. The semantic "agents" in series III constructions, and "experiencers" in inverted verbs, have less volitional control over the action than their counterparts in series I and II. And the narrative-marked "agents" in Pattern A are typically active (King 1994). So the constructions that make use of the different mapping patterns add more meaning and form restrictions than one would expect from just the mapping rules themselves. As the translations above demonstrate, series III forms have very strong evidential interpretations, in addition to their TAM properties.

It is also quite likely that there is more than one type of inversion, as suggested by Merlan (1982), Cherchi (1997), Tuite (p.c.). Harris' analysis of inversion relies on a purely syntactic notion of initial and final grammatical functions. As such, she claims that inversion in series III of active verbs, and inversion in indirect verbs are the same process.

However, a thematic role-based analysis must take into account further distinctions between the different types of inversion. While the dative-marked arguments in series III can still be called agents, the "deep" subjects of indirect verbs are more like experiencers, and their objects are more like stimuli. Thus, an inversion construction applicable for indirect verbs must make use of different thematic roles. In some sense, Merlan's suggestion that the inversion pattern is a separate argument-structure type in Georgian (Merlan 1982) is correct even in light of a mapping analysis of inversion. The morphosyntactic constructions in the following sections demonstrate how construction-specific restrictions combine with the choice of syntactic mapping pattern.

3.2.7 Summary of Issues

The data in this section indicate that Georgian has several distinct patterns of case-marking and possibly several patterns of agreement marking. The case-marking patterns are associated with different thematic roles depending on conjugation class and series. Agreement markers also map onto thematic roles differently depending on the conjugation class and series of the verb form. However, the distinctions made by the agreement morphology are different from those made by case-marking: whereas case-marking assignment depends on series and conjugation class, agreement marking seems to depend only on surface grammatical relations (which, in turn, are related to series and conj class). There is no direct correspondence between case markers and agreement markers; knowing what case is used to mark a particular argument will not necessarily predict what agreement set will be used to index its properties on the verb.

Case markers also cannot be used to distinguish between obligatory and optional arguments. The single argument of a verb can be marked by the Nominative, Dative, or Narrative case. However, the case-marking and agreement patterns do map onto each other quite consistently. While they may seem irregular viewed in isolation, regularity emerges when these patterns are viewed at the intersection of series and conjugation classes.

To summarize the morphosyntactic regularities, here is an informal constructional representation of a series.

(60)
$$\left[\begin{array}{l} \textit{construction-Series} \\ \textsc{sem} \quad \left[\textsc{Series TAM semantics}\right] \end{array}\right]$$

(61)
$$\left[\begin{array}{ll} \textit{construction-Series-ConjugationClass} & \\ \textsc{sem} & \left[\textsc{Series TAM semantics}\right] \\ \textsc{syn} & \left[\begin{array}{l}\textsc{Argument thematic roles and linking}\\ \textsc{Argument case-marking}\\ \textsc{Argument number specification}\\ \textsc{Argument person specification}\end{array}\right] \\ \textsc{morph} & \left[\begin{array}{l}\textsc{Morphological TAM properties}\\ \textsc{Verbal agreement properties}\end{array}\right] \end{array}\right]$$

The number agreement and other processes sensitive to more than 'surface' grammatical relations confirm, in addition to different linking constraints, that separate constructions are needed to represent the morphosyntax of series III inversion and indirect verbs. I now turn to the internal constitution of Georgian verb forms, which also turns out to have significant regularities determined by screeves and series.

3.3 Morphotactic Patterns

3.3.1 Structure of the Verb Template

A striking characteristic of Georgian verb forms is that the same formal elements (morphs) can recur in different formations, but it is almost impossible to assign consistent one-to-one correspondences between these formal elements and elements of meaning. Georgian verb forms contain many indentifiable morphs, which are sometimes taken to be a sign of *agglutinative* structure (Hewitt 1995b, Hillery 2001). However, as this section aims to demonstrate, many parts of Georgian verb formation are best classified as *flexive* and, to a certain extent, *templatic* in the sense of Bickel and Nichols (2006).

The structure of a Georgian verb roughly follows the template in (62). Not all of the parts of the template show up in all forms of a verb, but when they do show up, the order is always the same. These are rough divisions, as it is not always possible to cleanly divide a given verb form into these slots.

(62) Georgian verb template, from (Hewitt 1995b)

1. Preverb I (PVI), e.g. *uk'u-* 'back'

2. Preverb I (PVII), e.g. *she-* 'into'
3. Preverb III (PVIII), e.g. *mo-* 'towards the speaker'
4. Prefixal pronominal marker (PRON1), e.g. *v-* '1sg'
5. Pre-radical vowel (PRV), e.g. *i-*, *a-*, *e-*, *u-* indicates beneficiaries of actions and a host of other functions
6. Root, e.g. *xat'* 'draw'
7. Passive marker (PASS) *-d*
8. Thematic suffix (TS), e.g. *-av*, *-eb*
9. Causative formant (CAUS) *(-ev)-in*
10. 'Extension marker' (EM), e.g. *-d*, has no independent meaning
11. Screeve (tense) marker (SCR), e.g. *-i*
12. Suffixal person marker (PRON2), e.g. *-s* '3sg', *-en* '3pl'

Slots in linear order:

(PVI)-(PVII)-(PVIII)-(Pron1)-(PRV)-root-(Pass)-(TS)-(Caus)-(EM)-(Scr)-(Pron2)

Some examples (following Boeder 2005) are given in (63 - 67).

(63) *uk'u-she-mo-g-a-brun-a*
PVI-PVII-PVIII-PRON1-PRV-turn-SCR/PRON2
'(s)he made you turn back here (AORIST).'

(64) *mo-m-e-c-i-t*
PVI-PRON1-PRV-give-SCR-PRON2
'Give it back to me (addressee in the plural. AORIST).'

(65) *da-g-a-c'er-in-eb-d-a-t*
PVII-PRON1-PRV-write-CAUS-TS-EM-SCR-PRON2
'(s)he would make you(pl) write it (CONDITIONAL; used in the consequent clause).'

(66) *da-u-xat'-av-s*
PVII-PRV-draw-TS-PRON2
'(s)he will draw it for him/her (FUT)' or: '(s)he has drawn it (PERF)'

(67) *ga-u-cocxl-d-eb-od-nen*
PVII-PRV-alive-PASS-TS-EM-PRON2
'they would become alive for her (CONDITIONAL)' or: '(if) they became alive for her (SUBJUNCTIVE; used in the "if-clause")'

For our purposes, a simplified template in (68) is sufficient.

(68) Simplified verb template in linear order:
(PV_1)-$(Pron1_2)$-(PRV_3)-$root_4$-(TS_5)-(Scr_6)-$(Pron2_7)$

The fact that only combinations of these elements are distinctive is what determines the non-compositionality of Georgian verb morphology. For instance, there is no single morpheme that indicates the screeve. The general issue is that the only way to identify a particular form as belonging to a particular screeve is by knowing which combination of morphs was used, and that a different combination of forms would signal a different screeve. However, even the combination of formants used for a particular screeve is not always regular. This is illustrated by all the screeves of the verb *xat'va* 'to paint' in Table 3.18 (agreement properties are held constant). At the same time, verb forms by and large are unambiguous[6], and the number of distinct combinations of morphs seems to be sufficient to identify all of the distinct meanings.

Screeve	2SGSUBJ:3SGOBJ	
PRESENT	*xat'-av*	paint-TS 'you paint him'
IMPERFECT	*xat'-av-di*	paint-TS-SCR 'you were painting him'
PRES. SUBJUN.	*xat'-av-de*	paint-TS-SCR 'you would paint him'
FUTURE	*da-xat'-av*	PV-paint-TS 'you will paint him'
CONDITIONAL	*da-xat'-av-di*	PV-paint-TS-SCR'(if) you would paint him'
FUT. SUBJUN.	*da-xat'-av-de*	PV-paint-TS-SCR 'you should paint him'
AORIST	*da-xat'-e*	PV-paint-SCR 'you painted him'
OPTATIVE	*da-xat'-o*	PV-paint-SCR 'you are to paint him'
PERFECT	*da-g-i-xat'-av-s*	PV-PRON1-PRV-paint-TS-SCR 'you have drawn him'
PLUPERFECT	*da-g-e-xat'-a*	PV-PRON1-PRV-paint-SCR 'you should have drawn him'

Table 3.18: All screeves of the verb *xat'va* 'paint'

The following sections examine several elements in the Georgian verb template and their interactions and suggest that, once lexically determined, the behavior of these elements is governed by screeves and series. In addition, the behavior of elements like the

[6]except for several types of structural (non-lexical) homonymies, e.g. between the Pluperfect Conjugation 1 and relative Conjugation 2 forms; Aorist Conjugation 1 forms with the pre-radical vowel *i-* and Conjugation 2 Aorist forms, and several others (Aronson 1984)

preverbs, thematic suffixes, and screeve endings makes it possible to divide Georgian verbs into inflectional classes similar to the familiar Indo-European classes, where a few leading forms predict the rest of a verb's paradigm (Matthews 1991).

3.3.2 Preverbs

Preverbs form a closed class of 9, most often with a (C)V structure. Most verbs have a preverb lexically associated with them, although there is also a group of verbs that do not have preverbs. The 9 modern preverbs are given in Table 3.19 along with their original spatial and deictic meanings. As in Slavic, the preverbs are historically spatial and deictic markers which have become grammaticalized and often mark aspectual distinctions which have nothing to do with space. Forms with preverbs are usually perfective, while preverb-less forms are usually imperfective.

mi-	'hither'
mo-	'thither'
a(gamma)-	'up'
ga(n)-	'out'
she-	'in'
cha-	'down into'
ga(r)da-	'across, through'
c'a(r)-	'away'
da-	'down'

Table 3.19: Modern Georgian Preverbs (from Hewitt 1995:148)

In series I (most saliently for Conjugation 1 verbs, but also fairly common in other conjugations), screeves without preverbs are Present, Imperfect, and Present Subjunctive. Adding preverbs to these forms produces forms of Future, Conditional, and Future Subjunctive, respectively. The preverbed forms in most cases differ from their preverb-less counterparts in aspect. However, some preverbed verb forms do not have preverb-less counterparts, and vice versa. Some preverbed forms are not perfective, e.g. *hqavs* 'has' ∼ ***mi****-hqavs* 'brings' (Goletiani 1970:214).

Aspectual differences alone do not account for the differences in tense meanings. Examples in (69), (70), and (71) demonstrate the present and future screeves which differ

by a preverb. Screeves in general have multiple uses, particularly in combinations with other screeves. (71c) exemplifies a conditional sentence which uses the Present Subjunctive in the protasis and the Conditional in the apodosis.

(69) a. *k'ac-i* *surat-s* *xat'av-s*
man-NOM picture-DAT paint.3SGSUBJ.PRES
'The man paints / is painting a picture.' (Present)

b. *k'ac-i* *surat-s* *da-xat'av-s*
man-NOM picture-DAT paint.3SGSUBJ.FUT
'The man will paint a picture.' (Future)

(70) a. *k'ac-i* *surat-s* *xat'av-da*
man-NOM picture-DAT paint.3SGSUBJ.IMPERF
'The man was painting a picture.' (Imperfect)

b. *k'ac-i* *surat-s* *da-xat'av-da*
man-NOM picture-DAT paint.3SGSUBJ.COND
'The man would (often) paint a picture.' (Conditional)

(71) a. *k'ac-i* *surat-s* *unda* *xat'av-de-s*
man-NOM picture-DAT should paint.3SGSUBJ.PRESSUBJ
'The man should be painting a picture.' (Present Subjunctive)

b. *k'ac-i* *surat-s* *net'av* *da-xat'av-de-s*
man-NOM picture-DAT wish paint.3SGSUBJ.FUTSUBJ
'The man wishes he would paint a picture.' (Future Subjunctive)

c. *k'ac-i* *rom* *xat'avdes,* *surat-s* *da-xat'av-da*
man-NOM if draw.PRESSUBJ picture-DAT paint.3SGSUBJ.COND
'If the man could paint, he would paint a picture.' (Present Subjunctive, Conditional)

In general, the aspectual difference alone is not enough to account for the differences in meaning and use between the screeves that are distinguished by preverbs. The preverbs are also used in all Series II and all Series III tenses. Most of these forms do not have preverb-less equivalents, though Tschenkéli (1958:158) discusses imperfective aorists as well.

There is no consensus regarding whether preverbs are inflectional or derivational elements. In cases where they distinguish present-tense from future-tense forms, they are clearly inflectional. However, the same root can appear with more than one preverb, with

the combinations meaning very different things (e.g. *c'era* "to write", *da-c'era* "to write", *mi-c'era* "to ascribe", *ga-c'era* "to impose a tax"), in which case the preverb seems to play a derivational function. In the cases where preverbs are derivational, the placement of agreement prefixes between the preverb and the stem would seem to violate the assumption that all derivation much be attached to the stem earlier (and thus be closer to the stem) than all inflection. And even in cases where preverbs are clearly inflectional, they change the meaning of the verb much more than the agreement markers, and their placement further from the stem would violate Bybee's Relevance hierarchy (Bybee 1985).

However, this is not a problem if one views agreement markers are fitting into predefined slots in the verb template, so that the distance from the root need not correlate precisely with order of attachment. See section 3.4 for a discussion of possible stem structure of Georgian verbs.

A sample of the different preverbs is illustrated in Table 3.20.

PRESENT	FUTURE
xat'av 'you draw him/her'	***da****-xat'av* 'you will draw'
akeb 'you praise him/her'	***she****-akeb* 'you will praise him/her'
atbob 'you warm it up'	***ga****-atbob* 'you will warm it up'
trav 'you decorate'	***mo****-rtav* 'you will decorate'
ik'et'eba 'it is getting locked'	***cha****-ik'et'eba* 'it will get locked'

Table 3.20: Verbs and Their Preverb Samples

3.3.3 Thematic Suffixes

Like preverbs, thematic suffixes (sometimes also called characteristic suffixes or stem formants) form a closed class of about eight and are lexically associated with verbs. The thematic suffixes currently in use, roughly in decreasing order of frequency, are *-eb*, *-ob*, *-av*, *-am*, *-ev*, *-en*, and *-i* (Hewitt 1995:143b). Thematic suffixes participate in screeve formation, as can be seen in Table 3.18. Thematic suffixes are more common and more numerous in Conjugation 1 verbs than elsewhere.

In addition to verb formation, thematic suffixes also serve as indicators of morphological class, often predicting the variation in other parts of screeve formation. This predictive power applies even in screeves where the thematic suffixes themselves are absent,

like the Aorist in Table 3.21. In the 2sgSubj forms of the Aorist, the screeve ending could be *-e* or -i, and the 3sgSubj ending could be *-a* or *-o*. The choice of endings can often (though not always) be predicted if one knows the thematic suffix associated with the particular verb. However, since the thematic suffix itself does not appear on the Aorist forms (*-e*, *i*, *-a*, and *-o* are not allomorphs or phonological variants of the thematic suffixes), the prediction of an ending takes place in the comparison between a present-tense and an aorist-tense form.

	SERIES I	SERIES II	
GLOSS	2SG PRESENT	2SG AORIST	3SG AORIST
'write'	*c'er-**ø***	*da-c'er-**e***	*da-c'er-**a***
'draw'	*xat'-**av***	*da-xat'-**e***	*da-xat'-**a***
'build	*ashen-**eb***	*a-ashen-**e***	*a-ashen-**a***
'destroy'	*angr-**ev***	*da-angr-**i-e***	*da-angr-**i-a***
'open'	*xsn-**i***	*ga-xsen-**i***	*ga-xsn-**a***
'condemn'	*gm-**ob***	*da-gm-**e***	*da-gm-**o***

Table 3.21: The relationship between thematic suffixes and Aorist screeve endings, Conjugation 1.

3.3.4 Pronominal Agreement Markers and Screeve Endings

Before examining the individual screeve endings, it will be useful to look at morphotactic interactions between them and agreement markers, which questions the rigidity of the slots in the Georgian verbal template. However, even here all behavior is consistent within a particular screeve.

The pronominal agreement markers attach to the verb stem in completely regular ways when the subject (or object in inversion) is 1st or 2nd person: if the object is plural, the suffix *-t* attaches to the singular-object form; if the singular-object form had the suffix *-s*, then the plural ending is just *-t*, without the *-s* (Table 3.22).

If the subject is 3pl, the interaction is also regular: each screeve has its own ending which is the same regardless of the properties of the object (Table 3.23).

If the subject is 3sg, the situation is less regular. As Table 3.24 demonstrates, in most cases a screeve-specific ending takes precedence. If, however, the object is 2nd person plural, the interaction between the screeve-specific ending and the plural marker *-t* is not

	Non-inverted patient or inverted agent (object)		
Screeve	2sg	2pl	3sg / 3pl
present	gxat'av	gxat'av-**t**	xat'av
imperfect	gxat'av-**di**	gxat'av-**di-t**	vxat'av-**di-t**
pres. subj.	gxat'av-**de**	gxat'av-**de-t**	vxat'av-**de-t**
aorist	dagxat'-**e**	dagxat'-**e-t**	davxat'-**e-t**
aor. subj.	dagxat'-**o**	dagxat'-**o-t**	davxat'-**o-t**
perfect	dagixat'-i-**var***	dagixat'-i-**var-t***	dauxat'-i-**var***
pluperfect	dagexat'-**e**	dagexat'-**e-t**	davexat'-**e**

* - prescriptively, the thematic suffix *-av* is expected in these forms, but in everyday speech it is replaced by *-i*.

Table 3.22: Screeve endings and agreement markers for *xat'va* 'paint', 1sg subject forms

	Non-inverted patient or inverted agent (object)				
Screeve	1sg	1pl	2sg	2pl	3sg / 3pl
present	mxat'av-**en**	gvxat'av-**en**	gxat'av-**en**	gxat'av-**en**	xat'av-**en**
imperfect	mxat'av-**dnen**	gvxat'av-**dnen**	gxat'av-**dnen**	gxat'av-**dnen**	xat'av-**dnen**
pres. subj.	mxat'av-**dnen**	gvxat'av-**dnen**	gxat'av-**dnen**	gxat'av-**dnen**	xat'av-**dnen**
aorist	damxat'-**es**	dagvxat'-**es**	dagxat'-**es**	dagxat'-**es**	daxat'-**es**
aor. subj.	damxat'-**on**	dagvxat'-**on**	dagxat'-**on**	dagxat'-**on**	daxat'-**on**

Table 3.23: Screeve endings and agreement markers for *xat'va* 'paint', 3pl subject forms

always the same. In some cases, such as the present 3sgSubj marker *-s*, only *-t* follows the stem. In other cases, such as the Imperfect 3sgSubj ending *-da*, both the screeve ending and the plural marker remain. This interaction between screeve endings and agreement markers is screeve-specific rather than verb-specific.

A schematic representation of the relationship between series and morphology is in (72) and (73).

(72) Semantic and syntactic restrictions look the same as in (61).

	Non-inverted patient or inverted agent				
Screeve	1sg	1pl	2sg	2pl	3sg / 3pl
present	mxat'av-**s**	gvxat'av-**s**	gxat'av-**s**	gxat'av-**t**	xat'av-**s**
imperfect	mxat'av-**da**	gvxat'av-**da**	gxat'av-**da**	gxat'av-**da-t**	xat'av-**da**
pres. subj.	mxat'av-**de-s**	gvxat'av-**de-s**	gxat'av-**de-s**	gxat'av-**de-t**	xat'av-**de-s**
aorist	damxat'-**a**	dagvxat'-**a**	dagxat'-**a**	dagxat'-**a-t**	daxat'-**a**
aor. subj.	damxat'-**o-s**	dagvxat'-**o-s**	dagxat'-**o-s**	dagxat'-**o-t**	daxat'-**o-s**
perfect	damixat'av-**s**	dagvixat'av-**s**	dagixat'av-**s**	dagixat'av-**t**	dauxat'av-**s**/-**t**
pluperfect	damexat'-**a**	dagvexat'-**a**	dagexat'-**a**	dagexat'-**a-t**	daexat'-**a**/-**a-t**

Table 3.24: Screeve endings and agreement markers for *xat'va* 'paint', 3sg subject forms

(73)
$$\left[\begin{array}{l} \textit{construction-Series-ConjugationClass} \\ \textsc{morph} \left[\begin{array}{l} \textsc{Presence vs. absence of a preverb} \\ \textsc{Presence vs. absence of a thematic suffix} \\ \textsc{Interaction of screeve endings with agreement markers} \end{array}\right] \end{array}\right]$$

$$\left[\begin{array}{l} \textit{construction-Screeve-ConjugationClass} \\ \textsc{morph} \left[\begin{array}{l} \textsc{Screeve endings} \\ \textsc{Any other screeve-formation constraints} \end{array}\right] \end{array}\right]$$

3.3.5 Screeve Endings

As demonstrated above, the prefixes and thematic suffixes in isolation do not provide enough information to identify the screeve of an inflected verb form. An additional difficulty is presented by many of the so-called screeve endings themselves. The formation of the Aorist of Conjugation 1 verbs is particularly complex and is often not predictable from any semantically-based generalizations, as Table 3.21 hints. The most common template for an Aorist form (excluding person/number agreement) is as in (74):

(74) Preverb + stem + Aorist exponent (+ Plural)

The aorist behaves in some respects like a basic stem, given that it is morphotactically minimal and that it is used for unmarked functions like the 2p imperative. The Aorist exponent depends on the person of the subject and the verb type. In particular, 1st

and 2nd person subjects are marked the same, but differently from the 3rd person subjects. Table 3.25 demonstrates the different combinations of Aorist exponents. Hillery (2001) thoroughly examines the formation of the Aorist and identifies several morphological and phonological patterns that can often determine the screeve endings, but admits that overall "the aorist form of a Georgian verb is not wholly predictable, and must be learned."

	Type a	Type b	Type c	Type d
1,2 Subj	*-i*	*-e*	*-i*	*-e*
3Subj	*-a*	*-a*	*-o*	*-o*
1,2 Subj example	*mok'ali*	*dac'ere*	*miaq'rdeni*	*mosp'e*
3 Subj example	*mok'la*	*dac'era*	*miaq'rdno*	*mosp'o*
gloss	'kill'	'write'	'lean against'	'destroy'

Examples from (Hillery 2001).

Table 3.25: Aorist exponent combinations

There is no obvious correlation between the choice of 1st and 2nd subject Aorist exponent and the choice of 3rd person Aorist exponent for a given verb, and no external properties can predict to which type a given verb will belong. Some verbs also undergo stem changes in some forms of the Aorist, and the exact nature of the changes can only be determined lexically, verb-by-verb.

All of the allomorphs exemplified above have one thing in common: there are no meaning differences that could predict which form would be used, and the choice is not phonologically conditioned. The composition of particular verb forms depends on the lexical class to which they belong; and the only plausible way of determining lexical class is by comparison with other examples of the same class. The information on verb formation is concentrated in *paradigmatic comparisons* between full words forms rather than in any individual morphemes. Thus, a morpheme-based analysis could come up with meanings which would distinguish the different exponents, but the meaning difference would be unmotivated. Alternatively, one could add a set of mutual selection restrictions on which stems co-occur with which affixes, but that would violate the central tenet of morphemic approaches, that the meaning of the morphemes alone is enough to account for their distribution. In addition, a morpheme-based analysis would have to store a preverb, theme vowel, and thematic suffix in association with each lexeme, creating more unmotivated meaning distinctions. The attempt to re-assemble full words from their component parts invariably

results in loss of information that would be retained if one used full forms and contrasts between them. (see Blevins 2006 for a more detailed discussion of the informational content of different verb forms in Georgian).

At the same time, there is psycholinguistic evidence that people are capable of identifying small units of form that recur in related words and associating them with families of such words, even when the form units are clearly not morphemic or meaningful on their own. Bergen (2004) suggests that *phonaesthemes*, non-morphemic sound-meaning pairings such as the English *gl-* in sound and vision words like *glimmer*, *glow*, or *glance*, can be primed by other words containing them, but only when these words also have related meanings. The priming effects appear to be due to a combination of phonological *and* semantic commonalities, similar to known priming effects for fully compositional morphemes. However, phonaesthemes are not morphemes, and the remaining parts of words in which they occur (like *-immer*, *-ow*, or *-ance* above) have no obvious individual meanings. Bergen concludes that

> [...] while compositionality may play a role in morphological organization, statistical prevalence in the lexicon of form-meaning pairings is also a sufficient criterion for those pairings to display morpheme-like behavior in language-processing tasks. As such, these results also lend support to the prediction made by noncompositional morphological models that the frequency, and not just the compositionality, of recurring form-meaning pairings is crucial to their mental representation. (Bergen 2004:291)

Such evidence suggests that requirements on morphological compositionality can be relaxed, and that recurring formal parts of words can, indeed, be learned as abstractions over sets of full words that include them.

Even grammatical descriptions of Georgian that appear to be morpheme-based (Hewitt 1995b, Aronson 1990) simply segment word forms, describing minimal elements of morphological form. With the possible exception of agreement, these descriptions do not attempt to provide general rules for *constructing* verb forms from these smaller units, but instead rely on the predictive power of complete verb forms in order to infer the rest of a verb's paradigm.

On the other hand, a choice of aorist exponent often coincides with specific choices in other screeves. An Aorist form often serves as a leading form that can predict variation elsewhere in the verbal paradigm, even though other forms are not phonologically derived

from the Aorist. For example, the Pluperfect forms of Conjugation 1 and 3 verbs are formed from the Aorist, as discussed in section 2.3.2 and repeated in Table 3.26.

FUTURE	PERFECT	AORIST	PLUPERFECT	GLOSS
damalavs	*daumalavs*	*damala*	*daemala*	'hide'
dac'ers	*dauc'eria*	*dac'era*	*daec'era*	'write'
gaacnob	*gaucvnia*	*gaac'no*	*gaec'no*	'introduce'

Table 3.26: Georgian Perfect Tenses, Conjugations 1 and 3. (Based on Hewitt 1995)

The form interdependence is different, and more striking, in Conjugation 2 (unaccusative) verbs. Here, the Perfect and Pluperfect forms are constructed differently depending on whether the verb has an indirect object (bipersonal, or relative) or not (monopersonal, or absolute). Monopersonal forms are based on the participle forms, while the bipersonal forms are based on the verbal noun (*masdar*), as in Table 3.27.

PARTICIPLE	MONOPERSONAL	MASDAR	BIPERSONAL	GLOSS
dabadebuli	*dabadebuliq'o*	*dabadeba*	*dabadeboda*	'be born'
damaluli	*damaluliq'o*	*damalva*	*damalvoda*	'be hidden'

Table 3.27: Georgian Perfect Tenses, Conjugation 2. (Based on Hewitt 1995)

A strictly morpheme-based account would have no way of representing such associations. A more detailed division into lexical classes is exemplified below. Interestingly, the behavior of many morphs is consistent within a single series. For example, the use of theme vowels, thematic suffixes, and preverbs is the same for all screeves in a given series. Thus, morphotactic patterns provide some additional evidence that the series is a natural organizational unit in the language.

3.3.6 Conjugation Classes and Other Lexical Classes

The four conjugation classes of Georgian are distinguished on the basis of formal morphological characteristics such as screeve formation. Semantic properties are not sufficient to predict a given verb's conjugation class. As such, they look very much like lexical inflectional classes in other languages. Similar to situations in Latin or Russian, no single morpheme of a verb can unambiguously identify its conjugation class; rather, a combination

of morphs in different verb forms determines the class.

In addition, there is morphological variation within conjugation classes, as illustrated by the formation of the Aorist above. Conjugation classes can be further subdivided based on patterns of screeve formation. Some of the variation in screeve formation is predicted by the thematic suffix of a verb; other variation falls into more fine-tuned classes. Melikishvili (2001), in a study of approximately 9,000 verbs, divides them into 64 morphological classes. It is possible that the divisions in her classification are too fine, and the number of classes may indeed be smaller. However, it is beyond the scope of this study to evaluate exactly how many lexical classes are necessary.

For those morphological patterns that are productive, however, the question remains: how should they be described, and how are novel words adapted to those patterns? Matthews (1972), in re-introducing a Word-and-Paradigm model for inflectional morphology, clarifies how a leading form of a paradigm can predict the rest of the paradigm. While the Georgian situation is complex in different ways than the Latin examples used by Matthews, it is still possible to provide a principal-parts analysis in which a few leading forms (generally 2-4) are sufficient for predicting the rest of the paradigm. In fact, traditional descriptions of Georgian such as (Tschenkéli 1958) are full of diagnostic forms, and even a morpheme-based description in (Hewitt 1995b) predicts variation based on properties of other forms (e.g. the preverb, or the behavior of the Aorist).

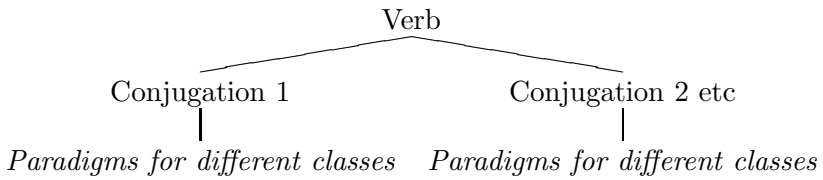
These diagnostic forms for Georgian are:

- Present form (predicts thematic suffix)
- Future form (predicts preverb and future-stem suppletion)
- Aorist 1st or 2nd person subject form (predicts the Aorist screeve ending)
- Aorist 3sgSubj form

For most verbs, suppletion is not an issue, and so the present form is unnecessary.

The division of verbs into lexical classes reflects recurring patterns in their formation. For each class, a certain set of forms can be used to predict the rest of that class's paradigm. In addition, verbs from the same conjugation/valency class share certain syntactic and semantic properties that could be represented in an inheritance hierarchy as in (75).

(75)



The purpose of such a hierarchy is to represent successive levels of abstraction, or pattern generality. Such a hierarchy is constructed bottom-up in language acquisition (similar to the 'abstractive' processees described by Blevins (2006) and discussed in Chapter 2). The inheritance relationships by themselves are not sufficient to represent the implicational structure of the lexicon. Instead, the hierarchy simply provides 'hooks' into appropriate example paradigms, and is anchored in real word forms.

3.3.7 Georgian Lexical Classes and Productivity

The number of lexical classes relates to the question of productivity. A pattern detected by a linguist is not necessarily salient in the mind of a speaker, who may instead memorize individual verbs. Tomasello's studies of item-based language acquisition (Dodson and Tomasello 1998, Lieven et al. 2003, Savage et al. 2003, Tomasello 2000, Tomasello 2003) suggest very strongly that children extend patterns based on words and phrases they hear often, not on abstract regular rules. Dąbrowska (2002) suggests on the example of Polish neuter case endings that some patterns are never learned, regardless of how regular they may be, because they are not sufficiently frequent.

A pilot questionnaire designed to test the productivity of Georgian verb formation has revealed that the same may be true for Georgian. I designed the pencil-and-paper questionnaire on the basis of transitive (Conjugation 1) verb formation patterns mentioned in (Hewitt 1995b). Overall, 21 patterns were identified based on thematic suffixes, root structure, and screeve endings in various screeves. The patterns are described in Tables 3.28 and 3.29. The parameters of variation are as follows:

- Thematic suffix
- Root type: with a vowel (RV) or without a vowel (RV). Vowel-less roots sometimes have an epenthetic vowel in some series II or Pluperfect, as indicated in the table.
- Present screeve 3PlSubj ending (Pr.3PlSubj) is *en* for most verbs, but *-an* for a small group.
- Series II thematic suffix behavior (SII.TS): in most cases, thematic suffixes disappear, but some verbs retain a remnant.

- Aorist 1st or 2nd person subject exponent (Aor. 12Subj) can be *e* or *i*, with *e* being more frequent.
- Aorist 3rd person subject exponent (Aor. 3Subj) can be *a* or *o*, with *a* being more frequent.
- Optative exponent (Opt exp) can be *a* or *o*, with *o* being more frequent.
- Perfect screeve thematic suffix behavior (Perf.TS): some thematic suffixes disappear, but some stay
- Pluperfect exponent (PluP exp): in most cases *a* or *o*, following the choice of the Aorist 3p exponent, but *ia* for one group of verbs.

The less frequent choices are in bold face in the table.

In the spring of 2003, the questionnaire was distributed to adult speakers in Tbilisi, Georgia. Speakers were given sentences with nonce verbs in the forms judged to be informative according to the patterns above. Speakers were then asked to produce the novel verbs in other inflected forms which should make clear whether or not the new verb was placed in the appropriate pattern. An example of a question is in (76).

(76) Example question for pattern 1.
dalis q'vavilis ***dashuxeba*** *uq'vars.*
'Dali likes to **dashuxeba** flowers.'
gushin man 3 q'vavili _____.
'Yesterday she _____ 3 flowers (expected aorist: da(a)shuxa).'
gushin shen 2 q'vavili _____.
'Yesterday you _____ 2 flowers (expected aorist: da(a)shuxe).'
gogim ramdenime q'vavili unda _____.
'Gogi must _____ many flowers (expected optative: da(a)shuxos).'
gushin gogis q'vavilebi unda _____ *magram ar* _____.
'Yesterday Gogi should have _____ flowers (exp.: pluperfect daeshuxebi(n)a) but did not _____ (expected: perfect daushuxebia).'

The goal of the questionnaire was to determine qualitative tendencies in nonce verb production, not to get quantitatively precise results. There were only two questions designed to test each pattern, and so the results cannot be judged statistically significant. However, some instructive general patterns emerged and are described in Table 3.30.

Table 3.28: Transitive verb patterns, based on Hewitt 1995.

<table>
<tr><th>TS</th><th>NO.</th><th>MELIK.</th><th>ROOT TYPE</th><th>PRES
3PLSUBJ</th><th>SER II
TS</th><th>AOR
12SUBJ</th><th>AOR
3SGSUBJ</th><th>OPT
EXP</th><th>PERF
TS</th><th>PERF
EXP</th><th>PLUP
EXP</th></tr>
<tr><td rowspan="5"></td><td>1</td><td>28, 29</td><td>root V</td><td rowspan="5">en</td><td rowspan="5">ø</td><td rowspan="2">e</td><td>a</td><td rowspan="2">o</td><td>TS</td><td rowspan="5">ia</td><td>ia</td></tr>
<tr><td>2</td><td></td><td>no root V</td><td rowspan="3">o</td><td rowspan="4">ø</td><td rowspan="4">o</td></tr>
<tr><td>3</td><td>30</td><td>no root V
-l/r/n
AI.12.VI e</td><td rowspan="3">i</td><td rowspan="3">a</td></tr>
<tr><td>4</td><td rowspan="2"></td><td rowspan="2">no root V, -l/r/n,
AI.12.VI a</td></tr>
<tr><td>eb</td><td>5</td><td>-shv-</td><td>a</td><td>a</td></tr>
<tr><td>ø</td><td>6</td><td>18, 20</td><td></td><td>en</td><td>ø</td><td>e</td><td>a</td><td>o</td><td>ø</td><td>ia</td><td>a</td></tr>
<tr><td>en</td><td>7</td><td>21</td><td></td><td>en</td><td>en→ in</td><td>e</td><td>a</td><td>o</td><td>TS</td><td>ia</td><td>a</td></tr>
<tr><td rowspan="2">ev</td><td>8</td><td>22</td><td></td><td rowspan="2">en</td><td>ev→i</td><td>e</td><td rowspan="2">a</td><td>o</td><td rowspan="2">TS</td><td rowspan="2">ia</td><td rowspan="2">a</td></tr>
<tr><td>9</td><td></td><td>-rtm-, -rkm-,
m→v in SII,
Plup</td><td>ø</td><td>i</td><td>a</td></tr>
<tr><td rowspan="4"></td><td>10</td><td>19, 26</td><td>root V</td><td rowspan="4">en</td><td rowspan="4">ø</td><td>e</td><td rowspan="3">a</td><td>o</td><td rowspan="4">TS</td><td rowspan="4">s</td><td rowspan="3">a</td></tr>
<tr><td>11</td><td></td><td>no root V</td><td rowspan="2">i</td><td rowspan="2">a</td></tr>
<tr><td>12</td><td>27</td><td>no root V, -l/r/n,
AI.12.VI a</td></tr>
<tr><td>av</td><td>13</td><td></td><td>-rt-, roots with
variation</td><td>e</td><td>o</td><td>o</td><td>o</td></tr>
</table>

Table 3.29: Transitive verb patterns continued.

TS	NO.	MELIK.	ROOT TYPE	PRES 3PLSUBJ	SER II TS	AOR 12SUBJ	AOR 3SGSUBJ	OPT EXP	PERF TS	PERF EXP	PLUP EXP
	14	25	root V	**an**	ø	e	a	o	TS	ia	a
	15	24	no root V, SII.VI a, Plup.VI a								
i	16	23	some no root V, -r/n, AI.12.VI e			**i**		**a**			
	17	31, 32		en	ø	e	**o**	o	ø	ia	o
	18		some no root V, AI.12.VI e			**i**					
ob	19		some no root V, AI.12.VI a								
am	20	34		en	ø	**i**	a	**a**	TS	**s**	a
op	21			en	ø	**i**	**o**	o	ø	ia	o

PAT	TS	ROOT	COMMENTS
1	eb	RV	Very productive. Over 77 % of expected forms were correct, and in some instances as high as 100%
2	eb	No RV	Endings that are the same as in Pat 1 were correct most of the time; endings that were different were correct about half the time
3, 4	eb	No RV epenthesis	The rare Aorist and Optative endings were given about 20 % of the time; the rest conformed to the more frequent pattern
7	en	any	Suffix was not changed in series II most of the time
8	ev	any	Suffix was not changed in series II
9	ev	final *m*	Correct 'rare' Aorist ending often for one nonce word, rarely for another word. Thematic suffix often not recognized.
10	av	RV	Almost 100% correct aorist (frequent pattern); around 30% chose the unusual Perfective ending
11	av	No RV	Almost all conform to the more frequent patterns
12	av	No RV epenthesis	Almost all conform to the more frequent pattern; thematic suffix not recognized
14	i	RV	Only about 10% chose the rare 3Pl Present/Future ending; about 70% did not recognize thematic suffix
15	i	No RV epenthesis	About 60% correct (frequent pattern)
16	i	No RV final *r/n* epenthesis	About 60% correct (frequent pattern)
17	ob	any	Almost all conform to the frequent Aorist pattern
18-19	ob	No RV epenthesis	About 40% chose the rare Aorist pattern; the rest did not separate thematic suffix
20	am	any	Thematic suffix not recognized
21	op	any	Thematic suffix recognized for one of the two words, but endings were still regularized.

Table 3.30: Patterns resulting from the morphology questionnaire

Overall, it seems the patterns identified based on grammars are not equally productive. In particular, some of the thematic suffixes (*-en*, *am*, *op*) are no longer identified as separate elements that can be extended to novel items. The application of analogy seems to depend on the relative frequency of patterns and not just similarity to the novel word. Also, some of the patterns above appear to be closed-class, or require phonological similarity in addition to morphological similarity in order to extend analogically to a new item. Much further study is needed to decide which patterns and sub-patterns in Georgian are truly extendable, and under what conditions.

3.4 Summary of Issues

There are crucial lexical dependencies between the roots, preverbs, thematic suffixes, and screeve formants in Georgian. As a result, no individual root (or extended stem) provides sufficient information for the formation of full word forms. But there does seem to be a certain structure / order to how word forms are constructed. The order of attachment, determined by a how much a given element is lexically affiliated with the whole word, appears to be as in (77).

(77) Verb Stem Structure root > preverb | thematic suffx > version markers | passive suffix > screeve-specific endings > regular agreement

The Georgian verbal system is characterized by recurrent formal parts that do not have consistent associations to meanings in all contexts. Rather, the combination of parts, when viewed in contrast to other combinations, can signal the meaning of the whole word. Even in specific contexts, the association between morphs and meanings is often not one-to-one, i.e. not compositional.

Morpheme-based accounts are not well suited for representing such systems. The same form element would have to receive different meanings in different contexts (e.g. the agreement morphs), and the generalization of single form would be lost. In addition, a morpheme-based account would have to store many elements for each lexeme (root, preverb, thematic suffix, etc.) thus losing the generalization that the meaning of a word can be independently composed from the meaning of its parts.

It is not clear that this is the kind of problem that exponence rules offer a solution to. Stump (2001) and Anderson (1984, 1992) consider only verbal agreement—the

outermost, and most regular, layer of inflection. Even here, the proposed accounts apply to a small subset of verbal paradigms, and do not scale up to the full system of agreement. More importantly, no attempt has been made to provide a realizational account of the full stem structure of Georgian verb forms, let alone a morpheme-based description of more than a small fragment of the verbal system. One potential difficulty is that a rule-based account would need to divide verbal forms into parts, each of which realizes a distinct property. Such divisions are not always possible because conjugation class markers interact with agreement properties. A rule-based account would have to view such morphs as realizing both agreement properties and conjugation class, thus missing the recurrent morphs that realize only agreement properties.

On the other hand, viewing the morph distribution patterns at the level of the series makes them appear much more consistent, and makes the interaction between lexical and syntactic properties much easier to describe. The next section presents a more formal constructional analysis of the morphosyntactic patterns.

3.5 Constructional Analysis

3.5.1 The Series as a Construction

As demonstrated above, the series serves as a natural unit of organization in Georgian. It has its own particular formal, syntactic, and semantic properties and determines properties of its parts: linking, case on nominal arguments and agreement and morph selection on the verb. Thus, the series is precisely the type of object that Construction Grammar is ideally suited to represent: a combination of parts whose meaning is more than the sum of the meanings of the parts.

Each series construction is a template containing a verb and its nominal arguments. The construction specifies the TAM properties of the series itself, e.g. Evidential Past for series III. Series constructions are subdivided by conjugation class. Conjugation-specific series constructions specify the following properties:

- Case-marking and the linking between scene participants like Agent and Patient and grammatical relations like Subject and Object.
- The behavior of preverbs, version vowels, and thematic suffixes

Another subdivision of series constructions is by screeve. Each screeve construction

specifies its own TAM properties in addition to those of the series. In addition, conjugation class-specific screeve constructions specify restrictions on exponents for the relevant verb forms.

Finally, there is a set of conjugation-specific screeve constructions that inherit both from the conjugation-specific series constructions and from the screeve constructions, unifying their properties. This multiple inheritance hierarchy allows us to keep screeve-particular TAM properties separate so that the generalization is not lost in conjugation class-specific constructions.

3.5.2 Constructional Representations

The generic Clause construction has a set of syntactic roles (e.g. subject and object) and a set of semantic roles (e.g. agent and patient). It is essentially a very general type of scene with basic participants, where no specific information about the participants is specified (78). Optional constituents and bindings are in parentheses.

(78)
$$\begin{bmatrix} \textit{Construction-Clause} \\ \text{SEM} \quad \begin{bmatrix} \text{FRAME} \quad \text{“Action”} \end{bmatrix} \\ \text{SYN} \quad \begin{bmatrix} \text{ARG-ST} \\ \text{DET} \\ \text{COMPS} \end{bmatrix} \end{bmatrix}$$

A series I construction is in (79). (80) exemplifies a conjugation class-specific series construction, and (81) exemplifies a screeve construction.

(79)
$$\begin{bmatrix} \textit{construction-SeriesI} \\ \text{SEM} \quad \begin{bmatrix} \text{Series I TAM semantics} \end{bmatrix} \end{bmatrix}$$

(80)
$$\begin{bmatrix} \textit{construction-SeriesI-Conjugation1} & \\ \text{SEM} & \begin{bmatrix} \text{AGENT} & \boxed{1} \\ \text{PATIENT} & \boxed{2} \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \text{ARG-ST} & \langle \text{SUBJ } \boxed{1}, \text{ DIR-OBJ } \boxed{2} \rangle \\ \text{DET} & \boxed{1} \begin{bmatrix} \text{CASE} & \text{NOM} \end{bmatrix} \\ \text{COMPS} & \boxed{2} \begin{bmatrix} \text{CASE} & \text{DAT} \end{bmatrix} \end{bmatrix} \\ \text{MORPH} & \begin{bmatrix} \text{NON-INVERTED} & \\ \text{SUBJ} & \begin{bmatrix} \text{PERSON} & \boxed{1}.\text{PERSON} \\ \text{NUMBER} & \boxed{2}.\text{NUMBER} \end{bmatrix} \\ \text{DIR-OBJ} & \begin{bmatrix} \text{PERSON} & \boxed{2}.\text{PERSON} \\ \text{NUMBER} & \boxed{2}.\text{NUMBER} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The linking between thematic roles and syntactic arguments in the constructions is done by coindexing. A more visual representation is in Tables 3.15 and 3.16 above.

(81)
$$\begin{bmatrix} \textit{construction-screeve-Present} & \\ \text{SEM} & \begin{bmatrix} \text{TENSE} & \text{PRESENT} \\ \text{ASPECT} & \text{IMPERFECTIVE} \\ \text{MOOD} & \text{DECLARATIVE} \end{bmatrix} \end{bmatrix}$$

One possible representation is in Figure 3.1. At a zoomed-in level, the paradigm representations correspond to fully specified constructions, similar to the representations of Russian paradigms in Figure 2.1, Chapter 2.

Figure 3.2 represents a (partial) view of how generalizations are abstracted from existing paradigms along multiple dimensions. This diagram can be thought of as a multiple-inheritance diagram or a network, in which the information propagates from the stored exemplars up into the more general patterns. The verbs listed at the bottom represent *example paradigms* of some verb classes. The hierarchy represents the extraction of series, screeve, and conjugation class-specific patterns. It should be noted, however, that other patterns may well be extracted on the basis of the stored examples, similar to the Russian

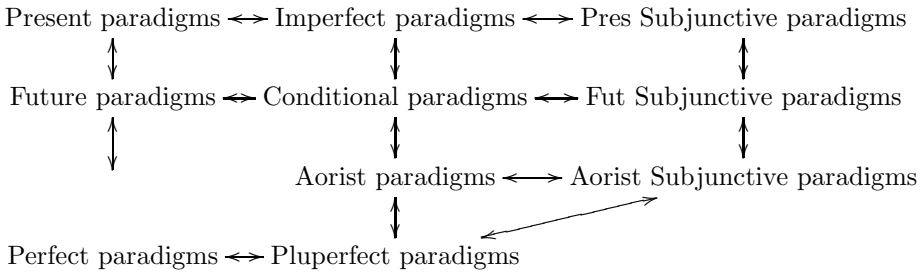


Figure 3.1: Network representation of Georgian paradigms

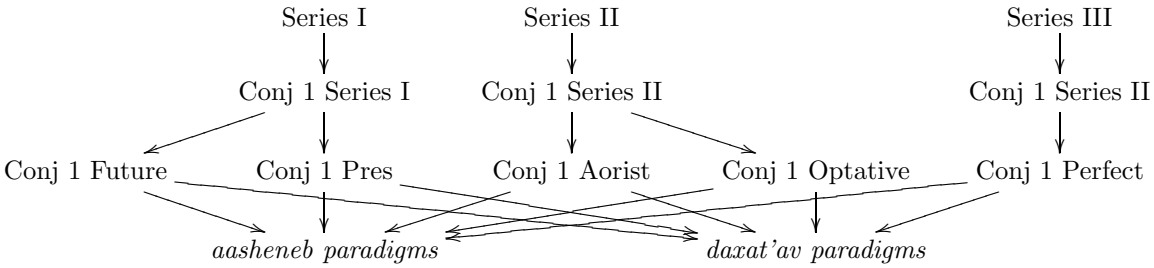


Figure 3.2: Partial Hierarchy of constructions

stress patterns in Chapter 2. The hierarchy differs from the standard HPSG or CxG constructional hierarchies in (at least) two ways. First, paradigms are explicitly represented as the grounding unit of organization. Paradigmatic patterns (such as rules of referral) can be extracted along with constructions that concern individual words. Second, the abstract constructions are exactly that - abstract, and are secondary to the stored exemplars.

The main purpose of the hierarchy as described above is for describing how novel forms can be inflected. Starting from the top of the hierarchy (with conjugation class and series information), a new word would pick up parts of a constructional representation as it moves down the hierarchy. Once the full set of morphosyntactic features is assembled, the word-formation process is presented with several example word for realizing these properties morphotactically. At this point, the process of analogy as described in the previous chapter is invoked, and produces one or more candidate formations, depending on the known forms of the novel word. Thus, morphosyntactic well-formedness of new words is assured by the successful path through the lattice, with the position of generalizations in the lattice corresponding to the order of application of constructions; and morphotactic well-formedness is assured by the process of analogy.

(82)
$$\left[\begin{array}{ll} \textit{construction-screeve-Present-conjugation1} & \\ \text{SEM} & \left[\begin{array}{ll} \text{TENSE} & \text{PRESENT} \\ \text{ASPECT} & \text{IMPERFECTIVE} \\ \text{MOOD} & \text{DECLARATIVE} \end{array}\right] \\ \text{SYN} & \left[\text{INHERITED FROM SERIES I-CONJUGATION 1}\right] \\ \text{MORPH} & \left[\text{INHERITED FORM SERIES I-CONJUGATION 1}\right] \\ \text{PHON} & \left[\text{AGR1 + ROOT + TS + AGR2}\right] \end{array}\right]$$

Recognition and production of new forms is done on the basis of the network of stored exemplars and patterns extracted from them, as in Figure 3.2. As emphasized before, the production of novel items takes place analogically, so the goal in production is to find an existing form maximally comparable to the form one wants to produce. Presumably, production starts with a desired set of morphosyntactic properties, and something known about the desired word (e.g. another inflected form of the same lexeme). Proceeding from the top of the hierarchy, the production process can 'pick up' morphosyntactic properties expressed by the increasingly more specific constructions, as long as they are compatible with the desired set. Eventually, this process will terminate in one or more stored forms, from which the new form is then created by analogy.

3.5.3 Summary

There are several advantages of the constructional analysis sketched out above. First, it resolves all mismatches by different mappings between the semantic roles and syntactic arguments. This way, the syntactic relations are kept within a single level, and no derivation between levels is required.

Second, the different mappings between syntax and semantics do not concern the morphology. Thus, agreement markers have a consistent association with syntactic arguments and there is only one pattern of morphological agreement. At the same time, the sensitivity of plural agreement to thematic roles rather than syntactic arguments can be naturally accommodated by constructional representations.

Third, the ability of constructions to refer to form and meaning properties of their constituents allows us to represent clause-level properties such as case marking in the same

constructions as verb-level properties such as agreement.

3.6 Discussion

The Georgian conjugational and agreement system has sometimes been described as chaotic and irregular. Yet, as I hope to have demonstrated, many of the seeming irregularities can be accounted for by a combination of lexical-class and constructional analysis. The long-term stability of the Georgian system demonstrates that it is not unlearnable and is not in need of simplification. The apparent mismatches between morphological and syntactic treatment of scene participants can be resolved by a few simple linking rules. Construction Grammar allows these rules to be represented in a single-layer, non-derivational format, so that these rules can interact with more idiosyncratic patterns without loss of their generality. A constructional representation allows syntax and morphology to be seamlessly integrated and represented in the same format.

A constructional analysis of linking issues makes predictions that are somewhat different from those made by derivational accounts based on general rules or transformations. In particular, a construction can easily become restricted to a particular class of lexemes or expand to accommodate novel instances. Thus, certain linking patterns can easily be applicable only to a certain class of verb, or verb forms.

The potentially narrow-class or even lexical nature of constructions fits well with evidence of item-based language acquisition. Imedadze and Tuite (1992) provide the most complete overview of studies on Georgian language acquisition to date. Their findings in general support the 'verb island' idea (Dodson and Tomasello 1998, Tomasello 2006). In particular, morphosyntactic patterns seem to be learned at first relative to specific words and groups of verbs: atelic verbs are first used in the present, and then in other Series I forms, following the Nominative-Accusative pattern; telic verbs are first used in the Imperative and Aorist, following the Narr-Nom pattern. In addition, the Narr-Nom pattern is often used with volitional subjects regardless of aspect, confirming at least for acquisition that an active pattern is present in Georgian rather than an ergative pattern. The two disjoint patterns are extended to other verb types only later.

Georgian children learn agreement and case-marking patterns for individual conjugation classes and series first, sometimes not extending the generalizations until they are 4 or 5 years old. Thus, at least for children, generalizations about agreement marking and

case-marking start at the level of individual verbs and series. The acquisition findings also support the word-based morphology hypothesis: children overextend frequent patterns but never produce forms which are not complete words.

The different phenomena present in Georgian — multiple-argument agreement, massive allomorphy, lexical classes, inversion, split between a Nominative/Accusative and Active patterns, slot competition — are not all that unusual typologically. What makes Georgian unique is the particular combination of morphosyntactic phenomena. As such, the language is a meta-example of a construction, where the whole is more than the sum of the parts.

The following two chapters examine in more detail one of the more complex 'parts' of the Georgian whole, the pre-radical vowels. On the one hand, pre-radical vowels have an easily identifiable places within the Georgian verbal template, and their morphotactic properties are quite clear. On the other hand, they interact in rather non-obvious ways with the larger constructions involving screeves, series, and conjugation classes. Pre-radical vowels are thus a perfect example of how complex form-meaning relationships can be. At the same time, pre-radical vowels follow the generalizations described earlier in this chapter, such as the series being a key organizational unit.

Chapter 4

Georgian Version

4.1 Introduction

In linguistic analysis, elements with a similar distribution are often taken to be similarly uniform in their syntactic properties. When applied to morphological structures, this type of assumption leads to analyses in which elements occupying a common 'slot' are treated as parts of a common morphological system. An examination of the phenomenon of *version*[1] in Georgian shows how heterogeneous the resulting systems can be.[2]

This chapter and the one that follows examine the function and distribution of the so-called *pre-radical vowels* (following Aronson's 1990 terminology) that fill the slot immediately preceding the verb root in the verbal template discussed in Section 3.3.1. The verb template is repeated in (83) for convenience.

(83) Simplified verb template in linear order:
(PV_1)-$(\mathrm{Pron1}_2)$-$(\mathbf{PRV}_3)$-root_4-(TS_5)-(Scr_6)-$(\mathrm{Pron2}_7)$

There are four possible pre-radical vowels: *i-*, *u-*, *e-*, and *a-*. I will argue that pre-radical vowels participate in a 'core' subsystem of version constructions that involve semantically-motivated alternations in argument structure, and a number of other functions that are merely marked in the same morphological position as core version markers. I present a detailed description and analysis of this system that brings out the different

[1]The term 'version' is a translation of the Georgian *kceva* meaning 'change,' and is described in the traditional grammatical literature as an inflectional category of a verb.

[2]An earlier version of the analysis presented here is presented in (Gurevich 2005b)

syntactic status and functions of the two fundamental groups of version vowels.

The core system of version contrasts is a semantic phenomenon with discourse-level repercussions, indicating a participant directly affected by an action; in the second, it has become grammaticalized in several distinct ways, interacting with the core morphosyntactic constructions of the language. In both cases, surrounding syntactic and semantic context determines the function of pre-radical vowels. Some of the constructions involving these vowels are connected by confirmed or plausible historical developments, while others appear to be outliers, unified with core version markers by a coincidence of identical distribution. Many uses of pre-radical vowels are historically and synchronically related, but it is not always clear to what extent a given usage of a pre-radical vowel is related to the meaning of version; the functions of the vowels form a continuum rather than a clearly divisible set. Thus, these markers will sometimes be called version vowels even if they do not synchronically seem to relate to valence-changing operations, if there are plausible historical connections to core version. However, to avoid making arbitrary decisions on when the use of such a vowel is version-related or not, they will always be glossed as PRV.

In the core 'version' subsystem, pre-radical vowels are used for a variety of valence- and thematic role- related functions, which are the focus of this chapter. Examples include valency-increasing marking of beneficiaries, reflexivity, marking of passives/impersonals, causatives, and the marking of indirect object agreement. One such use involving indirect-object agreement is in (84), where the pre-radical vowel changes depending on the person of the indirect object.

(84) a. *shen me k'oment'ar-s m-i-gzavn-i.*
2SG 1SG comment-DAT 1SGOBJ-PRV-SEND-SCR.
'You send me comments.' (Amiridze 2006:168)

b. *shen mas k'oment'ar-s u-gzavn-i.*
2SG 3SG.DAT comment-DAT PRV-SEND-SCR.
'You send him comments.' (Amiridze 2006:169)

In addition, version vowels have in some cases become grammaticalized and perform purely morphological functions, as discussed in Chapter 5. One example of a purely morphological use is the future-tense formation in (85).

(85) a. *pren / pren-s*
fly / fly-SCR
'You are flying / S(he) is flying (PRESENT)'

b. ***i***-*pren* / ***i***-*pren-s*
prv-fly / **prv**-fly-SCR
'You will fly / (S)he will fly (FUTURE)'

Version is traditionally described as one of the inflectional categories of the Georgian verb (Shanidze 1953, Melikishvili 2001, Hewitt 1995b). Aronson (1982) claims that it is more similar to a derivational category. As examples in this chapter will illustrate, version is somewhere in between, which makes it a perfect case study for the Constructional Morphology framework.

Pre-radical vowels occupy a specific morphotactic slot on the verb, just before the verb root; their distribution is thus uniform. Their function, however, varies depending on the tense / aspect, valency, and lexical class of the verb. No one pre-radical vowel can be given a uniform functional or semantic analysis. On the other hand, historical and distributional facts do indicate that most uses of pre-radical vowels are related morphologically. Version thus represents a mismatch between form and function that is not attributable to pure homonymy: although the functions of pre-radical vowels are largely distinct, their morphological and phonological behavior suggests an underlying commonality of form, and historical developments in many cases point to possible connections as well. In addition, many of these distinct functions are similar to the very diverse functions of reflexive clitics in Slavic and Romance, and more generally, to the functions of *middle voice* in Indo-European, as discussed in Chapter 7.

This chapter presents an analysis of version in the Constructional Morphology framework described earlier, unifying form and meaning and showing how version integrates with the rest of the Georgian morphosyntax. The function of pre-radical vowels is constant within the main organizational construction of Georgian, the intersection of conjugational class and TAM series described in the previous chapter. The present analysis of pre-radical vowels involves a comprehensive description of the function and distribution of these markers, not available in the existing literature, as well as a significant extension to the mechanisms of Construction Grammar in order to handle rich morphology and complex interactions between morphology, syntax, and semantics.

The main advantage of a constructional analysis is that it allows the description of regular patterns at various degrees of abstraction, both at the level of form (morphology) and meaning (cf. Booij (2005)). In the case of version vowels, constructions integrating contextual facts such as tense, valency, etc. allow one to describe the regular function of

version vowels in those restricted circumstances, without having to assume that the markers themselves carry regular meanings uniform across contexts. Essentially, constructions allow us to record commonalities between elements of form (somewhat) independently from elements of meaning, and vice versa.

I proceed now from semantically motivated constructions to morphological ones, in the process addressing the following issues:

- The distribution of version vowels seems to depend on other properties of the verbs with which they occur, and this relationship has not been fully explored (though see Boeder 1968 for the most comprehensive effort.)

- The function of version vowels in different contexts remains somewhat opaque. Version vowels appear to mark some very different grammatical properties in different morphosyntactic contexts. The historical and synchronic connections between different uses of version vowels have not been clearly established.

- In cases where version vowels appear to be optional, or where version-bearing verbs appear to be semantically interchangeable with non-version verb forms, the factors determining the choice have not been identified or described.

- The syntactic effects of version vowels have similarly not been examined or categorized. Harris (1981) provides a syntactic analysis for one productive use of version vowels, where version prompts the elevation of a noun phrase to indirect object status. However, the syntactic status of other version uses remains unclear. Similarly, the interaction between version vowels and other aspects of Georgian verbal morphology, such as agreement markers and object camouflage, require a deeper investigation.

I describe the related version uses as a *family* of related constructions. At the center of the family are the most semantically-motivated uses of version (e.g. objective version in transitive verbs), with successive branching into more lexicalized and more grammaticalized, less semantically determined, uses. This chapter focuses on the semantically motivated uses of pre-radical vowels (or 'version proper'), and the next chapter describes the morphological functions of pre-radical vowels that do not bear a synchronic connection to core version.

4.1.1 Version as Participant Affectedness

The data described in this chapter suggest that for many cases, the slot in which pre-radical vowels appear has a coherent general function. Quite often, the elements in this slot indicate changes in valence and grammatical voice. However, the syntactic arguments over which such elements take scope, and their precise function, are determined by the larger syntactic and semantic context in which the verb forms appear, as well as the lexical semantics of the verbs. The functions of the version elements can be further subdivided into several subsystems.

The most prototypical and most often described uses of version vowels involve the marking of *participant affectedness*, meaning that some discourse participant is directly affected by the action denoted by the verb. This term is related to the cognitive notion of *salience* or *prominence* (cf. Langacker 1991).

Version marking participant affectedness is most often seen on active verbs, where version vowels differentiate the largest number of types of affectedness. Here, objective version (affected participant is the indirect object), subjective version (subject), and locative version (indirect object) are possible. Version markers can also express participant affectedness on passive and impersonal verbs, but, as will be demonstrated below, the types of affectedness distinguished for these verbs are more restricted and less differentiated than for active verbs. In each of these cases, the function of version vowels crucially depends on the surrounding morphosyntactic contexts.

4.2 Active Verbs—Objective Version

The functions of various version vowels in the context of active verbs involve various departures from the default mapping between thematic roles and syntactic arguments. These functions apply only to the non-inverted contexts, i.e. series I and II of Conjugations 1 (transitive) and 3 (unergative) verbs. Using the linking convention introduced in the previous chapter, the default linking pattern can be represented as in Figure 4.1.

'Objective version' (known as *sasxviso* 'for the other' in traditional Georgian grammars) indicates that the event is directed towards and / or affects a participant which is most often coded as an indirect object. Objective version is the least lexically restricted type of version, and is most often optional. It is attested on transitive (Conjugation 1) verbs,

Intransitive (Conjugation 3)

agent	(**beneficiary/goal**)
↓	↓
subj	(**oblique**)

Transitive (Conjugation 1)

agent	*patient*	(**beneficiary/goal**)
↓	↓	↓
subj	*obj*	(**oblique**)

Ditransitive (Conjugation 1)

agent	*patient*	*recipient*	(**beneficiary/goal**)
↓	↓	↓	↓
subj	*obj*	$ind-obj$	**oblique**

Figure 4.1: Active Verbs: Default Mapping

and some active intransitive (Conjugation 3) verbs. Syntactically, objective version seems to be quite similar to applicative constructions described in the literature (Peterson 1999, Peterson forthcoming), and especially to applicatives in Bantu (Alsina and Mchombo 1990, Wunderlich 1991). However, the overall uses of version vowels are much more diverse than the typical applicative category, and the resulting version constructions have some unique semantic properties that are not expected from a fully productive lexical alternation like the applicative.

4.2.1 Objective Version in Transitive Verbs

If the indirect object is 1st or 2nd person, the version vowel is *i-*. In (86), the goal of the action (1sg pronoun) is expressed in a postpositional (oblique) phrase. By contrast, in (86b), the beneficiary is encoded as an indirect object; the verb agrees with it by means of the pronominal affix *m-*, and indicates the affectedness relationship with the version vowel *i-*. Recall from Chapter 3 that verbal agreement is expressed via a combination of a pre-stem marker (in this case *m-*) and a post-stem marker, which is null for 3rd person subjects.

(86) a. *meri-m* *da-xat'a* ***chem-tvis*** *surat-i*
Mary-NAR PV-paint.AOR **1sg-for** picture-NOM
'Mary painted a picture for me.[3]'

b. *meri-m* *da-**m-i**-xat'a* ***(me)*** *surat-i*
Mary-NAR PV-1SGOBJ-**prv**-paint **(1sg)** picture-NOM
'Mary painted a picture for me.'

If the indirect object is 3rd person, the version vowel is *u-*, as in (87). There is no indirect object marker on the verb. It has been suggested that *u-* is a fusion of 3rd person object marker *s/h-* and *i-*, but this is difficult to confirm independently because the phoneme *h-*, which would be the allomorphic variant of the suffix, is fairly rare and unstable in Modern Georgian.

(87) *meri-m* *da-**u**-xat'a* ***(mas)*** *surat-i*
Mary-NAR PV-**prv**-paint **(3sg)** picture-NOM
'Mary painted a picture for him/her.'

The syntactic effect of objective version is to elevate the affected participant to 'core argument' status. It is marked by the dative case and is cross-referenced on the verb by object markers. The direct object is often possessed by the indirect object, or becomes his/her possession. The indirect object is usually animate. The thematic role of the indirect object is often a recipient or beneficiary, different from the less restricted 'goal' thematic role of the oblique in a version-less variant. Semantically, the use of version signals a change in degree of affectedness of the indirect object, as discussed in Section 4.2.4.

Aside from case marking, another piece of evidence for the indirect-object status of the affected participant comes from *object camouflage*, a syntactic process described by Harris (1981). The Georgian verb has two agreement slots, with the prefixal slot usually occupied by object markers (as described in Chapter 3). 1st and 2nd person objects are signaled by overt person markers in this slot, but 3rd person objects are unmarked. In principle, the object slot can reference a direct or an indirect object. If there is no indirect object, the verb straightforwardly agrees with it (88a). If the verb has both a direct and indirect objects, and both objects are 3rd person, the verb can be said to agree with all three arguments, and neither argument triggers an overt object prefix (88b)[4]. If the indirect object is 1st or 2nd person, it triggers an object marker on the verb (88c).

[3]Unless otherwise indicated, all data are from native-speaker elicitation by the author.

[4]It can also be said that the indirect object triggers the zero allomorph of the IO prefix *s-/h-/ø-*.

Problems arise when the verb has both objects and the direct object is 1st or 2nd person. The direct object may not control object agreement on the verb, because the indirect object has priority (88d). Instead, in a process called 'object camouflage', the direct object is converted into a possessor phrase with the reflexive pronoun *tavi* 'head'. The resulting phrase is in 3rd person and does not require overt verb agreement (88e) and the 'conflict' is resolved. Object camouflage also applies if both the direct and indirect objects are 1st or 2nd person (88f)

(88) a. *vanom shen da-g-xat'-a*
Vano-NAR you PV-2OBJ-paint-SCR
'Vano painted you.'

b. *vanom anzori she-adar-a givis*
Vano-NAR Anzor-NOM PV-compare-SCR Givi-DAT
'Vano compared Anzor to Givi.' (Harris 1981:48)

c. *vanom anzori shen she-g-adar-a*
Vano-NAR Anzor-NOM you PV-compare-SCR
'Vano compared Anzor to you.' (Harris 1981:48)

d. * *vanom shen she-g-adar-a givis*
Vano-NAR you PV-2OBJ-compare-SCR Givi-DAT
'Vano compared you to Givi.' (Harris 1981:48)

e. *vanom sheni tavi she-adar-a givis*
Vano-NAR your head-NOM PV-compare-SCR Givi-DAT
'Vano compared you to Givi.' (Harris 1981:49)

f. *vanom chemi tavi shen she-g-adar-a*
Vano-NAR my head-NOM you PV-2OBJ-compare-SCR
'Vano compared me to you.'

Object camouflage applies in version forms when the direct object is 1st or 2nd person (89).

(89) *vin mo-m-i-k'l-a sheni tavi!*
who pv-1-ver-kill-aor your head.nom
'qui t'a tue pour moi!' ['Who killed you for me!'](Boeder 1968, citing (Vogt 1938:20))

In addition to the indirect-object status of the affected participant, there are discourse and semantic factors that distinguish the version constructions from their postpositional counterparts, as discussed in more detail in section 4.2.4.

4.2.2 Objective Version in Ditransitive Verbs

When objective version is used with a ditransitive verb, the syntactic effect is the promotion of beneficiary / maleficiary to indirect object. The object markers on the verb co-reference the same direct object referred to by version. The original indirect object is usually demoted to an oblique postpositional phrase, and it is nearly impossible to elicit sentences with two indirect objects. In (90b), the use of a version marker coincides with the indirect objecthood of the beneficiary / maleficiary, and the demotion of the original indirect object (stealee). However, in (91), the use of the version marker does not cause any demotions and instead results in the addition of a fourth argument to the verb, essentially creating two indirect objects. Pronouns are often dropped in Georgian, but the argument structure can be recovered from verb agreement.

(90) a. *man mo-**h**-p'ara (mas) vashl-i*
he.NAR PV-**3Obj**-steal.AOR (3SG.DAT) apple-NOM
'He stole an apple **from him/her**.' (Boeder 1968)

b. *man mo-**m**-**i**-p'ara mis-gan vashl-i*
he.NAR PV-**1sgObj-prv**-steal.AOR 3SG.GEN-from apple-NOM
'He stole an apple **for me** from him/her.'

(91) *gtxov, (me) gada-**m**-**i**-c-e es c'igni P'avles!*
please, (1SG) PV-**1sgObj-prv**-give-AOR this book.NOM Paul.DAT
'Please give this book to Paul for me!' (Boeder 1968)

This difference may be due to the impossibility of having two identically coded beneficiaries / maleficiaries. In (91) Paul is a *goal*, with a different thematic role from 'me', whereas in (90), both 'me' and 'him' would be beneficiaries/maleficiaries.

4.2.3 Objective Version in Intransitive Verbs

The uses of objective version with intransitive verbs are more restricted than with transitive verbs. The result of the action is always directed at the affected participant, but exactly how this is done is determined by the lexical semantics of the verb, and not all verbs can combine with objective version. In many cases, the impossibility of a version marker seems to be motivated by the lack of a semantic connection between the affected participant and the object / result of the verb's action.

In (92a), the objective version form of the verb means to perform the dancing action for someone; this is the most productive use of objective version consistent with transitive examples above.

(92) a. *shen m-**i**-cek'v-av (me)*
you 1SGOBJ-**prv**-dance-TS (1SG)
'You dance **for me**.'

b. *shen **u**-cek'v-av (mas)*
you **prv**-dance-TS (3SG.DAT)
'You dance **for him**.'

In (93b), the objective version verb form becomes transitive and idiomatic, meaning to make someone fly / go away. The verbs from Conjugation 2 typically form their future tense using the pre-radical vowel *i-* (see Chapter 5 on the relationship between the future marker *i-* and core version). So the sentence in (93b could potentially be ambiguous between an objective-version verb and a future-tense verb with an added direct object 'you'. However, the future marker *i-* does not vary with the person of the subject or object, whereas the objective version marker becomes *u-* with 3rd person indirect objects. As example (93c) demonstrates, the vowel really is marking objective version. In addition, direct objects of active verbs in series II are marked by the nominative case (see Chapter 3), whereas indirect objects are in the dative. As shown by (93d), the beneficiary/maleficiary of the flying action is indeed an indirect object.

(93) a. *v-pren / v-i-pren*
1SUBJ-fly / 1SUBJ-PRV-fly
'I fly / I will fly.'

b. *exla da-g-i-pren shen*
now PV-2OBJ-PRV-fly 1SG
'I will fly you away (=make you go away / throw you out).'

c. *exla da-u-pren mas*
now PV-PRV-fly 3SG.DAT
'I will fly you away (=make you go away / throw you out).'

d. *da-u-prine mas*
PV-PRV-fly.AOR 3SG.DAT
'I flew you away (=made you go away / threw you out; Aorist).'

In (94b), the objective version form of 'sing' is even more idiomatic, meaning to sing *along* with someone.

(94) a. *v-mγeri*
1SUBJ-sing
'I sing.'

b. *me mi-g-i-mγer-eb*
1sg pv-2-ver-sing-ts
'I will sing along with you.'

For other verbs, the only context in which intransitive Conjugation 1 verbs allow objective version is by adding an object which is somehow related to/possessed by the affected participant (95b), thus making the verb transitive.

(95) a. * *shen me m-i-cxovr-eb*
2SG 1SG 1SGOBJ-PRV-live-TS
'You (will) live for me.'

b. *am bich's tu k'argad m-i-cxovr-eb, me da-g-a-sachukreb*
this boy.DAT if well 1SGOBJ-PRV-live-TS, 1SG PV-2OBJ-PRV-reward-TS
'If you give this boy a good life for me, I will reward you.'

Finally, in some cases the use of version forms implies a possessed object even when it is not overtly present (96).

(96) *mo-m-i-mat'-eb*
PV-1OBJ-PRV-increase-TS
'you will increase for me (e.g., my salary)' ('increase' by itself is an intransitive / medioactive verb)

4.2.4 Discourse Factors Relevant for Objective Version

The indirect object is usually affected via some connection to the object or the result of the action. The object is very often inalienably possessed by the affected participant, as in (97); the interpretation could be benefactive or malefactive.

(97) *is g-i-t'ex-s shen mk'lavs*
he 2OBJ-PRV-break-3SG your arm.DAT
'He breaks (your) arm for you / on you.' (Boeder 1968)

The object can be in an animate relation to the affected participant like relatives or children (98, 99)

(98) *she-m-i-k-e* *me* *shvili*
PV-1SGOBJ-PRV-praise-AOR 1SG child.NOM
'You praised (my) child for me.'

(99) *is* *kali* *ga-m-i-gizh-eb-s* *bavshvebs*
this woman.NOM PV-1SGOBJ-PRV-crazy-TS-3SG child.PL.DAT
'This woman will drive (for me) (my) kids crazy.'

Finally, the object can simply be in possession of the affected participant. In such cases, the version form implies possession (100a); the *-tvis* form implies non-possession (100b). A non-version construction, if possible, has a different meaning from its version counterpart, and the meaning difference is motivated by the general meaning of "participant affectedness" carried by version.

(100) a. *is* *g-i-t'ex-s* *shen* *doks*
he 2OBJ-PRV-break-3SG your clay-jug.DAT
'He breaks a clay jug for you.' (Implication: your jug)

b. *is* *shentvis* *doks* *t'ex-s*
he you.for clay-jug.DAT break-3SG
'He breaks a clay jug for you.' (Implication: not your jug)

When there is a choice between using a version and a non-version construction, the general deciding factor seems to be the degree to which a participant is affected by the action. Greater degree of affectedness usually corresponds to use of version, whereas lesser degree of affectedness corresponds to a *-tvis* postpositional phrase or some other device. This choice works both ways: given that a version construction was used, the interpretation has to be that of more affectedness. In addition, the non-version constructions can sometimes be interpreted as involving either the more affected or the less affected construal, whereas the use of version unambiguously signals the more affected interpretation.

The more specific parameters affecting the choice are as follows, with the more affected variant on the left:

- Possession by indirect object vs. non-possession
- Inalienable possession vs. alienable possession
- Beneficiary of action vs. goal / recipient of action

In some cases, the postpositional and version constructions are equivalent and freely interchangeable. However, the version construction is preferred when the affected participant is directly linked to the action denoted by the verb, directly affected by it, and/or present at the event. The interpretation of (101a) is ambiguous between a recipient (more affected) and a beneficiary; however, the recipient interpretation sounds odd in (101b) and is impossible in (102b).

(101) a. *shen* ***m-i-****mγeri.*
2SG **1sg-prv**-sing.PRES.2SG
'You sing **for me / to me**.'

b. *shen* ***chem-tvis*** *mγeri.*
2SG **1sg-for** sing.PRES.2SG
'You sing **for me / ?to me**.'

The difference in affectedness is stronger for verbs whose semantic frames include an implicit affected participant.

(102) a. *is* ***me m-i-****q'viris.*
3SG **1sg 1sg-prv**-yell.PRES.3SG
'He yells **for me / at me**.'

b. *is* ***chem-tvis*** *q'viris.*
3SG **1sg-for** yell.PRES.3SG
'He yells **for me / *at me**.'

If the direct object of the verb is inalienably possessed by the affected participant, the version construction is the only one possible in (103a). In (103b), the postpositional construction is only possible if the direct object is not associated with / owned by the beneficiary.

(103) a. *chit'ma da-****m-i-****k'ort'na xeli*
bird.NAR PV-**1sg-prv**-peck hand.NOM
'The bird pecked me on the hand.'

b. *chit'ma* ***chem-tvis*** *da-k'ort'na* ****xeli / magida.***
bird.NAR **1sg-for** PV-peck ****hand** / **table**
'The bird pecked ***[my hand]/table for me**.'

The essential function of objective version is to encode the beneficiary of the action as an indirect object, thus changing the argument structure and the linking between

No version

agent	(*patient*)	(*recipient*)	(**beneficiary/goal**)
↓	↓	↓	↓
subj	(*obj*)	(*ind − obj*)	(**oblique**)

Objective Version

agent	(*patient*)	(*recipient*)	**beneficiary**
↓	↓	↙	↘
subj	(*obj*)	**ind-obj**	(*oblique*)

Figure 4.2: Objective Version

semantic and syntactic arguments. The use of objective version also narrows the possible range of the indirect object's thematic role. While the oblique argument in a *-tvis* phrase can be either a beneficiary or a goal, the indirect object in a version construction can only be a beneficiary (where the beneficiary could be negatively affected, as well). This is not entirely a surprise, since many thematic hierarchies (such as the one in (57) in the previous chapter) place beneficiaries higher than goals. As a simplified representation, consider the contrast between a postpositionally marked beneficiary/goal construction and an objective version construction in Figure 4.2.

In order to represent objective version constructionally, the following parameters need to be specified:

- The verb conjugation type (only Conjugation 1 or 3)
- The argument structure, with an extra indirect object
- The series of the verb form (only series I or II)
- The morphological changes

For simplicity, let us start with a specific construction, namely objective version in series I of conjugation 1 verbs.

(104)
$$\left[\begin{array}{ll} \textit{cxn-obj-version-conj1-seriesI} & \\ \text{SEM} & \left[\begin{array}{ll} \text{FRAME} & \text{AFFECTEDNESS} \left[\text{AFFECTED } \boxed{3}\right] \\ \text{AGENT} & \boxed{1} \\ \text{PATIENT} & \boxed{2} \\ \text{AFFECTED PARTY} & \boxed{3} \end{array}\right] \\ \text{ARG-ST} & < \text{NP } \boxed{1}, \text{NP } \boxed{2}, \text{NP } \boxed{3} > \\ \text{SYN} & \left[\begin{array}{ll} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & < \text{OBJ } \boxed{2}, \text{IND-OBJ } \boxed{3} > \end{array}\right] \\ \text{MORPH} & \left[(\text{PV}) + \text{AGR1} + \text{VER} + \text{ROOT} \ldots\right] \end{array}\right]$$

As a subcase specifying the version vowel *i-*, consider the following construction.

(105)
$$\left[\begin{array}{ll} \textit{cxn-obj-version-conj1-seriesI-12PIO} & \\ \text{SEM} & \left[\begin{array}{ll} \text{FRAME} & \text{AFFECTEDNESS} \left[\text{AFFECTED } \boxed{3}\right] \\ \text{AGENT} & \boxed{1} \\ \text{PATIENT} & \boxed{2} \\ \text{BENEFICIARY} & \boxed{3} \end{array}\right] \\ \text{ARG-ST} & < \text{NP } \boxed{1}, \text{NP } \boxed{2}, \text{NP } \boxed{3} > \\ \text{SYN} & \left[\begin{array}{ll} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & < \text{OBJ } \boxed{2}, \text{IND-OBJ } \boxed{3}\left[\text{PERSON } 1 \text{ OR } 2\right] > \end{array}\right] \\ \text{MORPH} & \left[(\text{PV}) + \text{AGR1} + /\text{i}/ + \text{ROOT} \ldots\right] \end{array}\right]$$

To show the interaction with a specific verb construction, consider the following constructions for the verb *xat'va* 'paint'. Without version (106), the verb frame requires an agent and a patient, which are co-referenced with the subject and object. An optional goal argument is encoded as an oblique. With objective version (107), the verb frame incorporates the meaning of affectedness, and the argument structure is changed to include an indirect object. Note that the optional goal argument is specified by the versionless construction, and not by the verb frame, so there is no need for two separate 'lexical entries' for *xat'va*.

(106)
$$\left[\begin{array}{ll} \textit{cxn-obj-xat'va-seriesI} & \\ \text{SEM} & \left[\begin{array}{ll} \text{FRAME} & \text{PAINTING} \\ & \left[\text{PAINTER} \quad \boxed{1}, \text{ PAINTEE } \boxed{2}\right] \\ \text{AGENT} & \boxed{1} \\ \text{PATIENT} & \boxed{2} \\ (\text{GOAL} & \boxed{3}) \end{array}\right] \\ \text{ARG-ST} & < \text{NP } \boxed{1}, \text{ NP } \boxed{2}, (\text{NP } \boxed{3}) > \\ \text{SYN} & \left[\begin{array}{ll} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & < \text{OBJ } \boxed{2}, (\text{OBLIQUE } \boxed{3}) > \end{array}\right] \\ \text{MORPH} & \left[(da) + \text{AGR1} + \text{VER} + \textit{xat'} \ldots\right] \end{array}\right]$$

(107)
$$\left[\begin{array}{ll} \textit{cxn-obj-xat'va-ObjVersion-seriesI} & \\ \text{SEM} & \left[\begin{array}{ll} \text{FRAME} & \text{PAINTING} \\ & \left[\text{PAINTER} \quad \boxed{1}, \text{ PAINTEE } \boxed{2}\right] \\ \text{FRAME} & \text{AFFECTEDNESS} \\ & \left[\text{AFFECTED} \quad \boxed{3}\right] \\ \text{AGENT} & \boxed{1} \\ \text{PATIENT} & \boxed{2} \\ \text{AFFECTED PARTY} & \boxed{3} \end{array}\right] \\ \text{ARG-ST} & < \text{NP } \boxed{1}, \text{ NP } \boxed{2}, \text{ NP } \boxed{3} > \\ \text{SYN} & \left[\begin{array}{ll} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & < \text{OBJ } \boxed{2}, \text{ IND-OBJ } \boxed{3} > \end{array}\right] \\ \text{MORPH} & \left[(da) + \text{AGR1} + \text{VER} + \textit{xat'} \ldots\right] \end{array}\right]$$

The version constructions combine with all of the regular constructions specifying TAM properties described in Chapter 3. Additionally, an indirect-object agreement construction is needed to specify the case and agreement properties (108). I am using the HPSG feature HEAD to represent the properties of the licensing verb; however, the notion

of 'head' is less restricted in many versions of Construction Grammar: the construction itself may stipulate which constituent serves as more primary, regardless of the intrinsic properties of that constituent[5].

(108)
$$\left[\begin{array}{ll} \textit{cxn-subj/ind-obj-agr} & \\ \text{ARG-ST} & \langle \ldots \text{ NP } \boxed{5} \rangle \\ \text{SYN} & \left[\begin{array}{ll} \text{HEAD} & \left[\text{AGR} \left[\begin{array}{ll} \text{SUBJ} & \left[\begin{array}{ll} \text{PERS} & \boxed{1} \\ \text{NUM} & \boxed{2} \end{array}\right] \\ \text{IND-OBJ} & \left[\begin{array}{ll} \text{PERS} & \boxed{3} \\ \text{NUM} & \boxed{4} \end{array}\right] \end{array}\right]\right] \\ \text{SUBJ} & \left[\begin{array}{ll} \text{PERSON} & \boxed{1} \\ \text{NUMBER} & \boxed{2} \end{array}\right] \\ \text{COMPS} & \langle \text{IND-OBJ } \boxed{5} \left[\begin{array}{ll} \text{CASE} & \text{DAT} \\ \text{PERSON} & \boxed{3} \\ \text{NUMBER} & \boxed{4} \end{array}\right] \rangle \end{array}\right] \\ \text{MORPH} & \left[(\text{PV}) + \text{AGR1} + \ldots + \text{AGR2}\right] \end{array}\right]$$

A more specific construction for 2sg subject / 1sg indirect object agreement:

[5]Dodge and Wright (2002), Wright and Kathol (2003) provide some arguments for constructional selection of head properties, particularly for purposes of agreement.

(109)
$$\left[\begin{array}{ll}
\textit{cxn-subj/ind-obj-agr-2SgSubj-1SgIO} & \\
\text{ARG-ST} & < \ldots \text{ NP } \boxed{1} > \\
\text{SYN} & \left[\begin{array}{ll}
\text{HEAD} & \left[\begin{array}{ll}
\text{AGR} & \left[\begin{array}{ll}
\text{SUBJ} & \left[\begin{array}{ll} \text{PERS} & 2 \\ \text{NUM} & \text{SG} \end{array}\right] \\
\text{IND-OBJ} & \left[\begin{array}{ll} \text{PERS} & 1 \\ \text{NUM} & \text{SG} \end{array}\right]
\end{array}\right]
\end{array}\right] \\
\text{SUBJ} & \left[\begin{array}{ll} \text{PERSON} & 2 \\ \text{NUMBER} & \text{SG} \end{array}\right] \\
\text{COMPS} & < \text{IND-OBJ } \boxed{5} \left[\begin{array}{ll} \text{CASE} & \text{DAT} \\ \text{PERSON} & 1 \\ \text{NUMBER} & \text{SG} \end{array}\right] >
\end{array}\right] \\
\text{MORPH} & \left[(\text{PV}) + /\text{m}/ + \ldots + \emptyset\right]
\end{array}\right]$$

4.3 Active Verbs—Subjective Version

Subjective version implies that the participant affected by the action is the subject. The meaning is reflexive, and the use of subjective version is much more restricted and less productive than that of objective version. In traditional descriptions of Georgian, subjective version is known as *sataviso* 'for oneself' or as an expression of possession. Vogt (1938:8) describes it as signalling "un rapport de possession entre le sujet et le régime direct, ou bien que l'action se fait au profit du sujet en faisant du régime direct la propriété du sujet[6]." As the examples below demonstrate, the semantic effect of subjective version is more varied than just literal possession.

The versionizer *i-* is always used, and there is no alternation with *u-* (110).

(110) a. *(me) saxl-s* *v-**i**-shen-eb*
(I) house.DAT 1SUBJ-**prv**-build-TS
'I build a house **for myself**.'

[6]a relationship of possession between the subject and the direct object, or that the action is done for the benefit of the subject in making the direct object the posession of the subject

b. *Meri* *saxl-s* ***i**-shen-eb-s*
Mary.NOM house.DAT **prv**-build-TS-3SGSUBJ
'Mary builds a house **for herself**.'

Subjective version is only possible with transitive (bivalent) verbs; semantic factors here seem to override the purely morphological distinctions between conjugation classes. Conjugation 1 verbs that are intransitive do not allow productive subjective version (111), and neither do intransitive Conjugation 3 verbs (112). Subjective version is not attested with ditransitive / trivalent verbs, and is not paraphrasable with a postpositional phrase.

(111) a. *v-a-mtknar-eb*
1SUBJ-PRV-yawn-TS
'I yawn.'

b. * *v-**i**-mtknar-eb*
1SUBJ-**prv**-yawn-TS
'I yawn for myself.'

(112) a. *v-cek'v-av*
1SGSUBJ-dance-TS
'I dance.'

b. * *v-**i**-cek'v-av*
1SGSUBJ-**prv**-dance-TS
'I dance for myself.'

Some uses of subjective version are lexically restricted, e.g. when the object marker and version vowel refer to different entities (113).

(113) *is* *gamo-**m**-**i**-dzax-ebs* *me*
he.NOM PV-**1sgObj-prv**-call-TS.PRES me
'He will call me **to himself**.' ambiguous with 'He$_i$ will call him$_j$ for me.' (Boeder 1968)

For some verbs, the vowel *i-* is always required. Some of these verbs may have semantics compatible with a subjective / reflexive meaning. In (114a) and (115), 'begin' and 'require' have a strong implication of self-affectedness. If a beneficiary is present (114b), it can be expressed in a postpositional phrase while the verb retains subjective-version marking. However, in some cases an objective version marker is also possible (114c), and

here the clause follows all the standard objective-version patterns such as indirect object agreement on the verb and case-marking on the nouns. Aside from common semantics of self-benefactivity, it is difficult to place these verbs in a more traditional classification.

(114) a. *me da-v-**i**-c'q'e /*davc'q'e lekcia*
1SG PV-1SUBJ-**prv**-start-AOR /*PV-start.AOR lecture.NOM
'I have started the lecture.' (frozen form) (Boeder 1968)

b. *me lekcia mis-tvis da-v-**i**-c'q'e*
1SG lecture.NOM 3SG.GEN-for PV-**prv**-start.AOR
'I have started the lecture **for him**.'

c. *me bavshv-s kartul-is sc'avleba da-v-**u**-c'q'e*
1SG child-DAT Georgian-GEN teaching PV-1SUBJ-**prv**-start.AOR
'I have started to teach the child Georgian (lit. I have started the teaching of Georgian for the child).'

(115) *roca sakme mo-i-tx-ov-s*
when work.NOM PV-PRV-require-TS-SCR
'When work will require.' (Internet)

Although productive subjective version is not possible in intransitive (unergative / Conjugation 3) verbs, some such verbs have lexicalized subjective version (116, 117, 118).

(116) *da-i-dzin-e*
PV-PRV-sleep-AOR
'You fell asleep (on purpose).'

(117) *i-cin-i*
PRV-laugh-TS
'You laugh.'

(118) *mo-i-mat'-eb*
PV-PRV-increase-TS
'You will gain weight (increase in size).'

There is no productive alternation between subjective-version constructions and versionless postpositionally marked reflexive constructions. Where a versionless construction is possible, it usually does not have a self-benefactive meaning (119b).

(119) a. *Meri saxl-s **i**-shen-eb-s*
Mary.NOM house.DAT **prv**-build-TS-3SGSUBJ
'Mary builds a house **for herself**.'

b. *Meri* *tavis-tvis* *saxl-s* ***a*-*shen*-*eb*-*s***
Mary.NOM self-for house.DAT **prv**-build-TS-3SGSUBJ
'Mary builds a house ***for herself/alone**.'

4.3.1 Subjective Version and Reflexivity

Amiridze (2006) identifies several reflexivization strategies in Georgian. Two of the subtypes involve the subjective version vowel *i*-. In the first, the vowel is the only thing necessary to signal co-reference between the agent and the beneficiary (120). These are the typical uses of subjective version as exemplified earlier in this section.

(120) Subjective version and reflexive 'indirect objects' (Amiridze 2006:159)

a. *me* *mo-v-i-ch'er-i* *namcxvar-i.*
1SG.NAR PV-1SGSUBJ-PRV-cut-SCR cake-NOM
'I cut myself [some] cake.'

b. *shen* *da-i-sx-i* *q'ava.*
2SG.NAR PV-PRV-pour-SCR coffee.NOM
'You poured yourself [some] coffee.'

c. *ana-m* *ga-i-tal-a* *vashl-i.*
ana.NAR PV-PRV-peel-SCR apple-NOM
'Ana peeled herself an apple.'

In the second strategy, the version vowel plus the body-part noun *tavi* indicate a reflexive relationship between the agent (subject) and a patient (direct object), as in (121). The reflexive examples confirm the syntactic generalization that subjective version can only be used when there is a formal direct object. These sentences also support, in a general sense, the notion that subjective version indicates a relationship of possession between the subject and the direct object.

(121) Subjective version and reflexive 'direct objects' (Amiridze 2006:98)

a. *me*$_i$ *v-i-k'rit'ik'-eb* *tav-s*$_i$.
1SG.NOM 1SGSUBJ-PRV-criticize-TS self-DAT
'I criticize myself.'

b. *shen*$_i$ *i-k'rit'ik'-eb* *tav-s*$_i$.
2SG.NOM PRV-criticize-TS self-DAT
'You criticize yourself.'

c. *is$_i$* *i-k'rit'ik'-eb-s* *tav-s$_i$.*
3SG.NOM PRV-criticize-TS-3SUBJ self-DAT
'He criticizes himself.'

The third reflexivization strategy involves a possessor-modified *tavi* phrase and is typically used without a subjective version vowel (122; the pre-radical vowel 'a' is a marker of lexical class known as 'neutral version' and will be discussed below. It does not mark subjective or objective version). The addition of subjective version vowels in these sentences results in a focused reading (123).

(122) Possessor-marked reflexivization (Amiridze 2006:97)

a. *me* *chem-i* *tav-i* *v-a-k-e.*
1SG.NOM my-NOM self-NOM 1SGSUBJ-PRV-praise-TS
'I praised myself.'

b. *shen* *shen-i* *tav-i* *a-dzag-e.*
2SG.NOM your-NOM self-NOM PRV-curse-TS
'You cursed yourself.'

c. *man* *tav-is-i* *tav-i* *ga-a-masxr-a.*
3SG.NOM self-POSS-NOM self-NOM PV-PRV-clown-TS
'He made fun of himself.'

(123) Possessor-marked reflexivization and subjective version (Amiridze 2006:99)

a. *me* *chem-i* *tav-i* *v-i-k-e.*
1SG.NOM my-NOM self-NOM 1SGSUBJ-PRV-praise-TS
'*I praised myself.'
'I praised MYSELF.'

b. *shen* *shen-i* *tav-i* *i-dzag-e.*
2SG.NOM your-NOM self-NOM PRV-curse-TS
'*You cursed yourself.'
'You cursed YOURSELF.'

c. *man* *tav-is-i* *tav-i* *ga-i-masxr-a.*
3SG.NOM self-POSS-NOM self-NOM PV-PRV-clown-TS
'*He made fun of himself.'
'He made fun of HIMSELF.'

Thus, subjective version is closely integrated with other reflexivization strategies in Georgian. However, as these examples demonstrate, it is not the only way to mark

reflexivity. And, as demonstrated by the lexicalized uses of subjective version, it does not always mark literal syntactic reflexivity. Thus, there is a complex interaction between subjective version and other construction types in Georgian morphosyntax.

Both objective and subjective version can be marked by the version vowel *i-*. However, their syntactic and semantic functions are quite different. Several attempts have been made to reconcile this difference. Mach'avariani (1987) suggests a unified semantic analysis, in which *i-* (used for 1st and 2nd person objective and all subjective version forms) indicates a 'centripetal', or introverted action, whereas the version marker *u-* (used to mark 3rd person objective version) indicates a 'centrifugal', or extroverted action. This explanation has intuitive appeal, although there do not appear to be obvious cross-linguistic parallels to this distinction.

Whatever the explanation, the use of *i-* in both types of version does not seem to be the result of pure homonymy, and its contribution to the meaning of the verb form is non-compositional. Rather, the presence of the vowel *i-* in an active verb form in conjunction with an indirect object and corresponding verb markers can be interpreted as objective version. On the other hand, the presence of *i-* in conjunction with a subject marker and no affected indirect object marks subjective version. In other words, the larger construction and the combination of other morphs determines the function of the version vowel.

Boeder (1968) suggests that subjective version is a subtype of objective version and adds a 'latent reflexive' to the verb's argument frame, which is co-indexed with the subject. Such an analysis makes sense given the reflexive functions described in the previous section, although the idea of adding an empty syntactic argument is somewhat counterintuitive, and seems even more questionable for verbs that have lexicalized subjective version. The same intuition can be expressed in a mapping approach by linking both the agent and the beneficiary to the grammatical subject, thus avoiding the introduction of an empty syntactic argument. An argument-linking representation of subjective version is in Figure 4.3, contrasted with a non-version linking. The patient/direct object argument is not optional in the case of subjective version.

A constructional representation of subjective version is as follows. The construction follows all of the regular subject/object agreement principles.

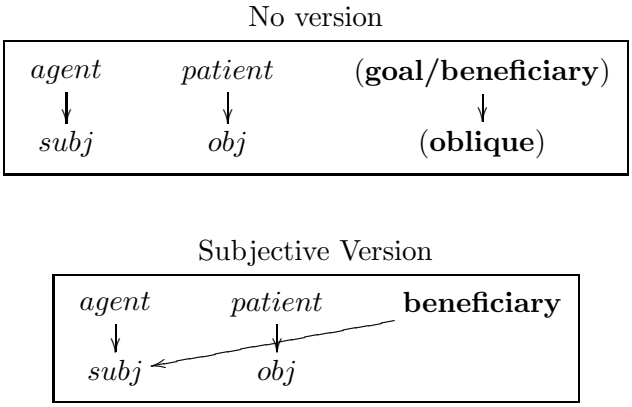


Figure 4.3: Subjective Version

(124)
$$\left[\begin{array}{ll} \textit{cxn-subj-version-conj1-seriesI} & \\ \text{SEM} & \left[\begin{array}{lll} \text{FRAME} & \text{AFFECTEDNESS} & \left[\text{AFFECTED } \boxed{1}\right] \\ \text{AGENT} & \boxed{1} & \\ \text{PATIENT} & \boxed{2} & \\ \text{BENEFICIARY} & \boxed{1} & \end{array}\right] \\ \text{ARG-ST} & < \text{NP } \boxed{1}, \text{NP } \boxed{2} > \\ \text{SYN} & \left[\begin{array}{ll} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & <\text{OBJ } \boxed{2} > \\ \text{SERIES} & \text{I} \end{array}\right] \\ \text{MORPH} & \left[\text{(PV) + AGR1 + /i/ + ROOT ...}\right] \end{array}\right]$$

4.4 Active Verbs—Locative Version

Locative version (also known as *superessive*, cf. Hewitt 1995b), indicates that the action was performed in some spatial relation to the indirect object. The spatial relation is usually 'onto', but metaphoric extensions are also possible. Locative version is marked by the pre-radical vowel *a-*. Shanidze (1953, 1980) considered locative version (or *sazedao* 'on something') to be part of a different subsystem called 'situation.'

In (125a), the ground of the breaking action (his head) is encoded as an indirect object, and the verb has the pre-radical vowel *a-*. The direct object (stick) is the figure that

moves relative to the ground. The same notion can be paraphrased with a postpositional phrase as in (125b).

(125) a. *me v-**a**-t'ex-av jox-s mis tav-s*
I 1SUBJ-**prv**-break-TS stick-DAT his head-DAT
'I break a stick **over** his head.'

b. *me v-t'ex-av jox-s mis tav-**ze***
I 1SUBJ-break-TS stick-DAT his head-**on**
'I break a stick **over** his head.'

Locative version is expressed the same way with stative verbs (126, 127).

(126) *a-k'er-ia*
ver-stitch-scr
'X is stitched onto Y.'

(127) a. *t'ot-s **a**-zis*
branch-DAT **prv**-sit.PRES
'He sits **on** a tree branch.'

b. *t'ot-**ze** zis*
branch-**on** sit.PRES
'He sits on a tree branch.'

Locative version appears to be restricted to transitive and stative verbs, but it is too rare for any definitive conclusions.

The spatial relation can be interpreted metaphorically, as in (128a) and (128b). While it is not entirely clear what the metaphorical extension is, it seems reasonable that some sort of spatial relation is implied in these examples between the subject (singer / yeller) and the object of singing / yelling.

(128) a. *(me) (shen) mo-g-**a**-mγere*
(1SG) (2SG) PV-2OBJ-**prv**-sing.AOR
'I sang you off / sang (at you) as you were leaving.'

b. *(me) (shen) mo-g-**a**-q'vire*
(1SG) (2SG) PV-2OBJ-**prv**-yell.AOR
'I yelled at you as you were leaving / yelled in your tracks.'

Some of the examples display a redundant expression of the spatial relation (129). The verb has the version vowel *a-*, but the ground is expressed by a postpositional phrase.

(129) *k'iser-**ze** **a**-zis*
neck-**on** **prv**-sit.PRES
'He sits **on (someone's) neck** / lives off of someone.' [7]

Another example of double-marking is given in (130). Here, native speaker intuitions indicate a difference in intentionality between the different syntactic encodings. Here, the locative-version construction is the least intentional, and the double-marked construction is the most intensional.

(130) a. *kvab-i* *da-v-dg-i* *navtkura-ze*
bowl-NOM PV-1SGSUBJ-put-SCR stove-ON
'I put a bowl on the stove.' (purpose unspecified: possibly in order to cook something)

b. *kvab-i* *da-v-a-dg-i* *navtkura-s*
bowl-NOM PV-1SGSUBJ-PRV-put-SCR stove-DAT
'I put a bowl on the stove.' (not with the intent of cooking)

c. *kvab-i* *da-v-a-dg-i* *navtkura-ze*
bowl-NOM PV-1SGSUBJ-PRV-put-SCR stove-ON
'I put a bowl on the stove.' (with the intent of cooking) (Example from Boeder 1968; interpretation elicited)

A similar distinction in intentionality applies in (131), where double marking is impossible. However, the matching of morphological expression to intentionality is different here, as the locative version-marked form appears to be *more* purposeful, although both the double-marked sentence in (130) and the locative-version sentence in (130) indicate greater affectedness of the indirect object. While it is clear that the use of locative version involves a difference in discourse / pragmatic interpretation, the precise parameters of variation need to be researched further.

(131) a. *c'gin-i* *da-v-d-e* *magida-ze*
book-NOM PV-1SGSUBJ-put-SCR table-ON
'I put a book on the table.' (purpose unspecified)

b. *c'gin-i* *da-v-a-d-e* *magida-s*
book-NOM PV-1SGSUBJ-PRV-put-SCR table-DAT
'I put a book on the table.' (with the intent to cover the table)

[7] Note that the expression 'to sit on someone's neck' could be a calque of Russian *sidet' na shee (u kogo-to)*, and so its syntax may be highly irregular.

In the cases of promotion of a goal/beneficiary to indirect argument, the resulting argument mapping has the same argument-linking properties as objective version (see Figure 4.2). But this change in argument linking is in some cases optional, as example 129 above demonstrates.

The difference in constructional representation is in the semantic frames involved. In addition to affectedness, locative version also expresses a spatial relation 'on' in which the affected participant (indirect object) is the ground. The figure is the patient, which is the object for transitive verbs (as in 132) and the subject for stative verbs.

(132)

$$\left[\begin{array}{ll}
\textit{cxn-loc-version-conj1-seriesI} & \\
\text{SEM} & \left[\begin{array}{lll}
\text{FRAME} & \text{AFFECTEDNESS} & \left[\text{AFFECTED} \quad \boxed{3}\right] \\
\text{FRAME} & \text{SPATIAL-RELATION} & \left[\begin{array}{ll}\text{RELATION} & \text{ON} \\ \text{FIGURE} & \boxed{2} \\ \text{GROUND} & \boxed{3}\end{array}\right] \\
\text{AGENT} & \boxed{1} & \\
\text{PATIENT} & \boxed{2} & \\
\text{BENEFICIARY} & \boxed{3} &
\end{array}\right] \\
\text{ARG-ST} & < \text{NP } \boxed{1}, \text{ NP } \boxed{2}, \text{ NP } \boxed{3} > \\
\text{SYN} & \left[\begin{array}{ll}
\text{HEAD} & \left[\text{SERIES} \quad \text{I}\right] \\
\text{SUBJ} & \boxed{1} \\
\text{COMPS} & < \text{OBJ } \boxed{2}, \text{ IND-OBJ } \boxed{3} > \\
\text{SERIES} & \text{I}
\end{array}\right] \\
\text{MORPH} & \left[(\text{PV}) + \text{AGR} + /\text{a}/ + \text{ROOT} \ldots\right]
\end{array}\right]$$

(133)
$$\begin{bmatrix} \textit{cxn-loc-version-stative} & & & \\ \text{SEM} & \begin{bmatrix} \text{FRAME} & \text{AFFECTEDNESS} & \begin{bmatrix} \text{AFFECTED} & \boxed{2} \end{bmatrix} \\ \text{FRAME} & \text{SPATIAL-RELATION} & \begin{bmatrix} \text{RELATION} & \text{ON} \\ \text{FIGURE} & \boxed{1} \\ \text{GROUND} & \boxed{2} \end{bmatrix} \\ \text{PATIENT} & \boxed{1} & \\ \text{BENEFICIARY} & \boxed{2} & \end{bmatrix} \\ \text{ARG-ST} & < \text{NP } \boxed{1}, \text{ NP } \boxed{2} > & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{SERIES} & \text{I} \end{bmatrix} \\ \text{SUBJ} & \boxed{1} \\ \text{COMPS} & < \text{IND-OBJ } \boxed{3} > \end{bmatrix} & \\ \text{MORPH} & \begin{bmatrix} (\text{PV}) + \text{AGR1} + /\text{a}/ + \text{ROOT} \ldots \end{bmatrix} & \end{bmatrix}$$

4.5 Version as Participant Affectedness, Non-Active Verbs

Version can also indicate participant affectedness and valency changes in some passive and impersonal verbs in Conjugation 2. However, the types of distinctions marked by version in Conjugation 2 are different from those marked in Conjugation 1. In particular, the types of affectedness are not differentiated to the same extent as in Conjugation 1. The main goal of this section is to demonstrate the difference between version in the two conjugation classes, and to argue that the larger contexts (conjugation class and series) again determine the function of version vowels.

Conjugation 2 includes many passive or impersonal counterparts of Conjugation 1 verbs. There are three ways to form Conjugation 2 verbs: by adding a pre-stem vowel (so-called prefixal passives), adding the suffix *-d* (suffixal passives), and without adding any extra elements. The two latter verb types follow the same pattern as active verbs with regards to version vowels.

For passives with prefixal formation, pre-radical vowels *i-* and *e-* can mark the presence (*e-*, 134b) or the absence (*i-*, 134a) of an affected participant, encoded as an indirect object. It is not known whether the pre-radical vowel *i-* used here is related to the

i- used in subjective or objective version, but see the discussion at the end of this section for possible connections.

(134) a. ***i**-xat'eba*
prv-draw.INTR.PRES
'It is being drawn.'

b. ***e**-xat'eba* *(mas)*
prv-draw.INTR.PRES (he.DAT)
'It is being drawn **for / in front of / on him/her**.'

The syntactic function of *e-* is to add an indirect object to the verb frame. Only the presence of an affected participant is marked; the way in which this participant is affected is either not specified or is provided by the lexical semantics of the verb. Further, subjective and locative versions are not possible. In a sense, the version vowel *e-* combines the functions of objective and locative versions.

Some such uses have become lexicalized, as in (135).

(135) a. *bich'-i* *pur-ze* *ga-**i**-gzavna*
boy-NOM bread-on PV-**prv**-send.AOR
'The boy got sent for bread.'

b. *bich'-i* *pur-ze* *ga-m-**e**-gzavna*
boy-NOM bread-on PV-1SGOBJ-**prv**-send.AOR
'The boy got sent for bread (by me), but he initiated the sending.'

Example (136) and (137) demonstrate that Conjugation 2 verbs are not always passive. Rather, this conjugation signals that the subject is not entirely in control, or the direct object is not prominent.

(136) ***i**-landzγeba*
PRV-curse.PASS
'(S)he is engaged in cursing' (not relevant who (s)he is cursing). (Nino Amiridze, p.c.)

(137) a. *shen she-i-jibr-eb-i*
2SG PV-PRV-compete-SCR
'You will compete (volitionally; you will initiate the action), you are worthy of competition.'

b. *shen ar* *she-i-jibr-eb-i*
2SG PV-PRV-compete-SCR
'You are hard to compete with.'

c. *shen she-e-jibr-ebi*
2SG pv-ver-compete-scr
'You will compete with X (initiate the competition in which someone else is also involved).'

Conjugation 2 verbs often become lexicalized as allowing only one or the other version marker. In (138b), the root meaning 'to use' plus version marker *e-* has become lexicalized as a bivalent verb 'to help'. This verb is no longer transparently connected to the verb 'use', and thus removing the version marker would take away its core meaning.

(138) a. *is shen g-e-xmar-eba*
3SG 2SG 2OBJ-PRV-help-SCR
'He helps you(sg).'

b. *is i-xmar-eb-a*
3SG PRV-help-SCR
'It is being used / (*it is being helped).'

Similarly, 'trust' in (139a) requires the version vowel *e-*.

(139) a. *Nino Vakhtangs mi-e-nd-oba*
Nino.nom Vakhtang.dat pv-ver-trust-scr
'Nino will trust Vakhtang.'

b. * *mi-i-nd-obi*
pv-ver-trust-scr
'trust'

When both vowels are possible, *e-* carries a very strong implication of an added affected participant, even when that participant is not overtly present in the sentence as an indirect object. In (140b), the implied second participant (1sg pronoun) is not present in the sentence. It does not trigger verbal agreement and therefore cannot simply be the result of pro-drop. On the other hand, the idiomatic use of this form does not preclude a more expected interpretation of a passive/impersonal with an added indirect object (140c).

(140) a. *she-i-k'vreb-i*
PV-PRV-group-SCR
'You will group yourself / concentrate.'

b. *she-e-k'vr-ebi*
PV-PRV-group-SCR
'You and I will group / do something "under the table" together.'

c. *shen p'olit'ik'ur p'art'ia-s she-e-k'vr-ebi*
2SG political party-DAT PV-PRV-group-SCR
'You will join / unite with a political party.' together.'

The affected participant interpretation is often blocked by more strongly lexicalized inceptive or patientive interpretations. In (141) the addition of the version marker *e-* is interpreted according to the productive pattern, meaning that the action affects some participant X. In (141) there is a strongly lexicalized interpretation of 'feeling like X' and something appearing to someone.

(141) a. *i-xat'-eba*
PRV-draw-SCR
'It is drawn.'

b. *e-xat'-eba*
PRV-draw-SCR
'It is drawn for him/her.'

c. *m-e-xat'-eba*
1SGOBJ-PRV-draw-SCR
'(*It is drawn for me) / I feel like drawing / It (an image) appears to me.'

In (142b) and (142c), the lexicalized meaning of 'register' is more entrenched, and the simple beneficiary interpretation of the form with *e-* is completely impossible. Due to zero-marking of 3rd person subjects, example (142b) can also be interpreted as '(s)he is registerting' where the 3rd person participant is the subject rather than object, and the object is not specified (cf. to 142d for a 1st-person subject form)).

(142) a. *i-c'er-eba*
PRV-draw-SCR
'It is written.'

b. *e-c'er-eba*
PRV-write-SCR
'(*It is written for him/her) / feels like writing / It will be written / (S)he is registering.'

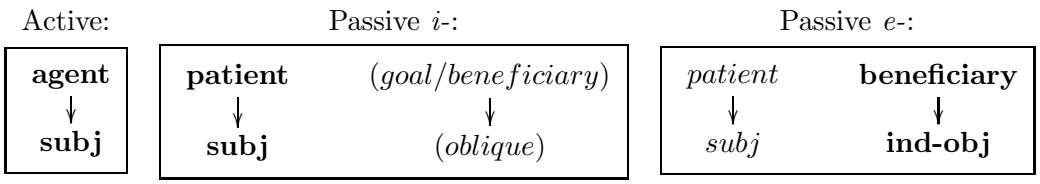


Figure 4.4: Passive / Impersonal Version

c. *m-e-c'er-eba*
1SGOBJ-PRV-write-SCR
'(*It is written for me) / I feel like writing / It is counted at my expense (I am accountable).'

d. *v-e-c'er-ebi*
1SUBJ-PRV-draw-SCR
'I am registering.'

Many Conjugation 2 verbs share roots with Conjugation 1 verbs. However, it is unclear how they came to have the vowels *i-* and *e-*. It is plausible that the use of *i-* is related to subjective version (reflexive) *i-* in Conjugation 1 verbs: it is quite common cross-linguistically for passive, impersonal, and reflexive verbs to have the same morphology (cf. Klaiman 1991, Blevins 2003, Israeli 1997). The use of the same morphological marker for directedness (as in objective version) and lack of volitionality (as in Conjugation 2 verbs) is also not unprecedented: a somewhat similar phenomenon is described for the Salish language family (Demirdache 1997). It has been suggested for Georgian that the passive marker *i-* developed from the subjective-version *i-* (Shanidze 1953, 1980, Imnaishvili 1968, Schmidt 1962, as quoted in Tuite 2003). However, there is no definite proof of this development, and it is conceivable that it might have gone the opposite direction (Alice Harris, p.c.).

The use of version vowels in passive / impersonal verbs signals an argument linking different from active verbs, although the use of a version vowel is not the only indication that a verb form is passive/impersonal. As a simplified representation of the linking contrast, consider Figure 4.4. The use of either *i-* or *e-* indicates that the subject is linked to the patient role. The difference between the two version vowels is in profiling: *i-* emphasizes the patient-subject linking, whereas *e-* emphasizes the addition of a beneficiary.

A constructional representation of passive/impersonal *i-* is given in (143).

(143)
$$\begin{bmatrix} \textit{cxn-conj2-prefixal} & \\ \text{SEM} & \begin{bmatrix} \text{PATIENT} & \boxed{1} \end{bmatrix} \\ \text{ARG-ST} & < \text{NP } \boxed{1} > \\ \text{SYN} & \begin{bmatrix} \text{SUBJ} & \boxed{1} \end{bmatrix} \\ \text{MORPH} & \begin{bmatrix} \text{(PV) + AGR1 + /i/ + ROOT ...} \end{bmatrix} \end{bmatrix}$$

A constructional representation of passive/impersonal *e-* is given in (144).

(144)
$$\begin{bmatrix} \textit{cxn-conj2-prefixal-ind-obj} & \\ \text{SEM} & \begin{bmatrix} \text{FRAME} & \text{AFFECTEDNESS} \begin{bmatrix} \text{AFFECTED} & \boxed{2} \end{bmatrix} \\ \text{PATIENT} & \boxed{1} \\ \text{BENEFICIARY} & \boxed{2} \end{bmatrix} \\ \text{ARG-ST} & < \text{NP } \boxed{1}, \text{ NP } \boxed{2} > \\ \text{SYN} & \begin{bmatrix} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & < \text{IND-OBJ } \boxed{2} > \end{bmatrix} \\ \text{MORPH} & \begin{bmatrix} \text{(PV) + AGR1 + /e/ + ROOT ...} \end{bmatrix} \end{bmatrix}$$

4.6 Causativization

The version vowel *a-* is used in forming causatives, sometimes accompanied by the thematic suffix *-eb* (Aronson 1991). For the most part, *a-* does not appear in forms derived from verbs such as the masdar.[8] Causatives formed from intransitive verbs are illustrated in (145).

(145) Intransitive causatives:

a. *v-q'ep*	b. *v-**a**-q'ep-**eb***
1SUBJ-bark	1SUBJ-**prv**-bark-**ts**
'I bark.'	'I make him bark.'

Causatives formed from transitive verbs may contain an extra suffix *-in*, as in (146b). As is consistent with cross-linguistic observations about causatives, it seems that

[8]There is some evidence of 'frozen' version vowels in some derived nouns (Tuite, p.c.), but finding these requires a more thorough etymological investigation than is possible within the scope of this dissertation.

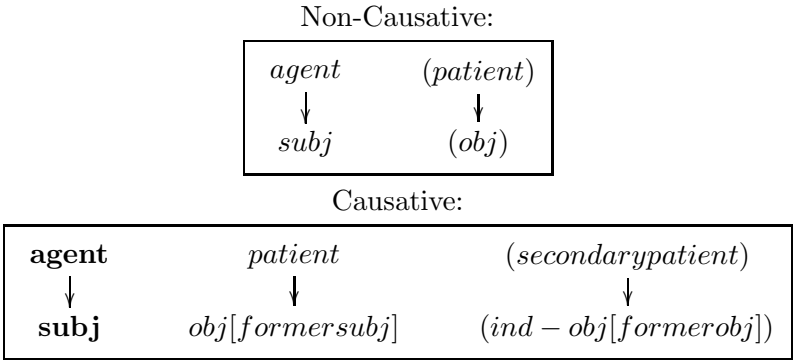


Figure 4.5: Causativization

the non-causative direct object ('letter') remains a direct object in a causative clause, while the non-causative subject ('he') becomes the indirect object and the non-causative indirect object is demoted to an oblique (cf. Harris 1981:67, Ackerman and Moore 1999).

(146) Transitive causatives:

a. *amxanag-s c'eril-s s-c'er-s*
friend-DAT letter-DAT *s/h*-write-3SGSUBJ3OBJ
'He writes a letter to a friend.'

b. *amxanag-is-tvis c'eril-s mas v-**a**-c'er-**in**-**eb** mas*
friend-GEN-for letter-DAT 3SG.DAT 1SUBJ-**prv**-write-**caus**-**ts**
'I make him write a letter to a friend.'

c. *amxanag-is-tvis c'eril-i mas da-v-**a**-c'er-**in**-e mas*
friend-GEN-for letter-NOM 3SG.DAT PV-1SUBJ-**prv**-write-**caus**-SCR
'I made him write a letter to a friend.' (Aorist)

Mach'avariani (1988) suggests a possible historical connection between the locative version marker *a-* and the causative *a-*, but this connection needs to be investigated further. Also, the same pattern as the causative, including the version vowel *a-* and the thematic suffix *-eb*, is also a very common way of forming verbs from adjective and noun stems and is discussed in more detail in Chapter 5.

The argument-linking representation of causatives is in Figure 4.5 and consists in the addition of an agent and the corresponding shift in argument linking, demoting the other arguments.

A constructional representation of a causative construction is in (147).

(147)
$$\left[\begin{array}{ll} \textit{cxn-causative} & \\ \text{SEM} & \left[\begin{array}{lll} \text{FRAME} & \text{CAUSATION} & \left[\begin{array}{ll} \text{CAUSER} & \boxed{1} \\ \text{CAUSEE} & \boxed{2} \\ (\text{OBJECT} & \boxed{3})\end{array}\right] \\ \text{AGENT} & \boxed{1} & \\ \text{PATIENT} & \boxed{2} & \\ (\text{SECONDARY PATIENT} & \boxed{3}) & \end{array}\right] \\ \text{ARG-ST} & < \text{NP } \boxed{1}, \text{ NP } \boxed{2}, (\text{NP } \boxed{3}) > \\ \text{SYN} & \left[\begin{array}{ll} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & < \text{OBJ } \boxed{2}, (\text{IND-OBJ } \boxed{3}) >\end{array}\right] \\ \text{MORPH} & \left[(\text{PV}) + \text{AGR1} + /\text{a}/ + \text{ROOT} \ldots\right] \end{array}\right]$$

4.7 Version Markers as Experiencer Agreement

4.7.1 *i/u-*

For many verbs, version markers have become grammaticalized as experiencer agreement. Such verbs subcategorize for an indirect object and may or may not have a direct object. The thematic role of the indirect object appears to be more general than a salient / affected participant. Rather, the indirect objects in these constructions seem to express general experiencers / undergoers. These constructions do not have version-less counterparts.

The verbs in (148) do not subcategorize for a direct object and belong to Conjugation 3.

(148) a. *m-i-chivl-eb*
1SGOBJ-PRV-sue-TS
'You will sue me.'

b. *m-i-chkmet'* *xelze*
1SGOBJ-PRV-pinch hand.on
'You will pinch me on the hand.'

In (149), the verb 'approve' is phrasal, including of a required direct object, and

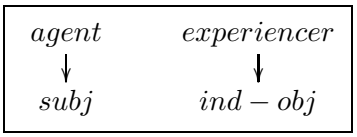


Figure 4.6: Experiencer Agreement

the patient is encoded as an indirect object and cross-referenced on the verb with a version vowel.

(149) *k'rebam* *da-**u**-ch'ir-a* *mxari* *am* *k'andidat'uris* *sheq'vanas*
council.NAR PV-**prv**-hold-SCR shoulder.NOM this candidacy.GEN inclusion.DAT
saarchevno *biulet'enshi.*
electoral ballot.in
'The council supported [lit: held a shoulder to] the inclusion of this candidacy in the electoral ballot.' (Internet)

The argument linking for experiencer-agreement constructions can be represented as in Figure 4.6. Constructionally, experiencer agreement via version vowels can be represented as in (150).

(150)
$$\left[\begin{array}{ll} \textit{cxn-experiencer-agreement} & \\ \text{SEM} & \left[\begin{array}{ll} \text{AGENT} & \boxed{1} \\ \text{EXPERIENCER} & \boxed{2} \end{array}\right] \\ \text{ARG-ST} & < \text{NP } \boxed{1}, \text{ NP } \boxed{2} > \\ \text{SYN} & \left[\begin{array}{ll} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & \left[\text{IND-OBJ} \quad \boxed{2}\right] \end{array}\right] \\ \text{MORPH} & \left[(\text{PV}) + \text{AGR1} + /\text{i/u}/ + \text{ROOT} \ldots\right] \end{array}\right]$$

The use of version vowels to express experiencer agreement may be an intermediate stage between between semantically / pragmatically motivated objective version and fully grammaticalized, morphological uses of version vowels.

4.7.2 *s/h-*

Object agreement in Georgian also sometimes includes the prefix *s/h-*, where the two consonants are phonologically conditioned allomorphs. The prefix cross-references 3rd

person indirect objects and cannot co-occur with a version vowel (151); the indirect object can fulfill a number of thematic roles.

(151) a. *k'ac-ma mona mepe-s mo-h-gvar-a*
man-NAR slave-NOM king-DAT PV-*s/h*-bring-SCR
'A man brought a slave to the king.'

b. *deda-m c'eril-i shvil-s mo-s-c'er-a*
mother-NAR letter-NOM child-DAT PV-*s/h*-write-SCR
'The mother wrote a letter to the child.'

Sometimes, constructions which use *s/h-* can be contrasted with other constructions in which the affected participant is expressed in a postpositional phrase, similar to the contrast between version vowel-containing constructions and those with other postpositional phrases (152).

(152) a. *kurd-ma shvil-i mama-s mo-s-t'ac-a*
thief-NAR child-NOM father-DAT PV-*s/h*-kidnap-SCR
'The thief stole a child from her father.'

b. *kurd-ma shvil-i mama-isa-gan mo-i-t'ac-a*
thief-NAR child-NOM father-GEN-from PV-PRV-kidnap-SCR
'The thief stole a child from her father.' (The pre-radical vowel *i-* is lexicalized subjective version)

Thus, in some ways, the constructions with *s/h-* indicate a degree of participant affectedness, can serve to elevate a participant to an indirect object (core argument) status, and are functionally somewhat similar to version markers. Such classification has been suggested by Shanidze (1926:337) and by Boeder (1969). Shanidze (1953, 1980) calls *s-/h-* marking *satanao* 'at/to someone' and suggests that it is another type of version. The forms with *s/h-* are not as common in modern Georgian as they once were (Braithwaite 1973:37), seem to be more lexically restricted than forms with version vowels, and their lexical distribution requires a more thorough investigation. However, it does seem that such forms, along with *i/u-* experiencer agreement forms, serve as the morphological basis for some of the inversion constructions discussed in the next chapter.

The argument-linking representation of *s/h-* forms is the same as that for *i/u-* experiencer agreement forms (Figure 4.6). The constructional representation is also similar, as in (153).

(153)
$$\begin{bmatrix} \textit{cxn-experiencer-agreement} & \\ \text{SEM} & \begin{bmatrix} \text{AGENT} & \boxed{1} \\ \text{EXPERIENCER} & \boxed{2} \end{bmatrix} \\ \text{ARG-ST} & < \text{NP } \boxed{1}, \text{ NP } \boxed{2} > \\ \text{SYN} & \begin{bmatrix} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & < \text{IND-OBJ} \quad \boxed{2} > \end{bmatrix} \\ \text{MORPH} & \begin{bmatrix} (\text{PV}) + \text{AGR1} + /\text{s/h}/ + \text{ROOT} \ldots \end{bmatrix} \end{bmatrix}$$

4.8 Discussion

The pre-stem slot on the verb is most often related to valency and voice. Most of the syntactic differences between different version constructions can be represented in terms of linking between thematic roles and syntactic arguments, as recapitulated in Figure 4.7. Objective and locative version signal that another argument is mapped onto the indirect object of the sentence. Moreover, objective version contrasts with version-less sentences in that the indirect object must be interpretable as a beneficiary / maleficiary, whereas in versionless sentences the indirect object position can also be occupied by a goal. Subjective version signals that the action involves a beneficiary which is co-referential with the agent / subject. The use of pre-radical *i-* with passive verbs signals, in addition to a change in transitivity, that subject is the sole undergoer / patient of the action. By contrast, *e-* in passive verbs refers to the presence of a beneficiary encoded as indirect object. A different transitivity-changing operation is the causative, which indicates the addition of another agent and subsequent demotion of the original agent to indirect objecthood. The use of *i-* and *u-* in experiencer agreement constructions is not paradigmatically opposed to a version-less construction and thus serves merely as indirect object agreement. Finally, a linking for inversion is given here for comparison, to indicate that it is a simple reversal in linking from the experiencer agreement construction. The use of pre-radical vowels in inversion is discussed in Chapter 5.

This generalization is not without exception. In some contexts, pre-stem vowels may have purely morphological functions, quite unrelated to syntactic valency or voice. Such cases are taken up in the next chapter.

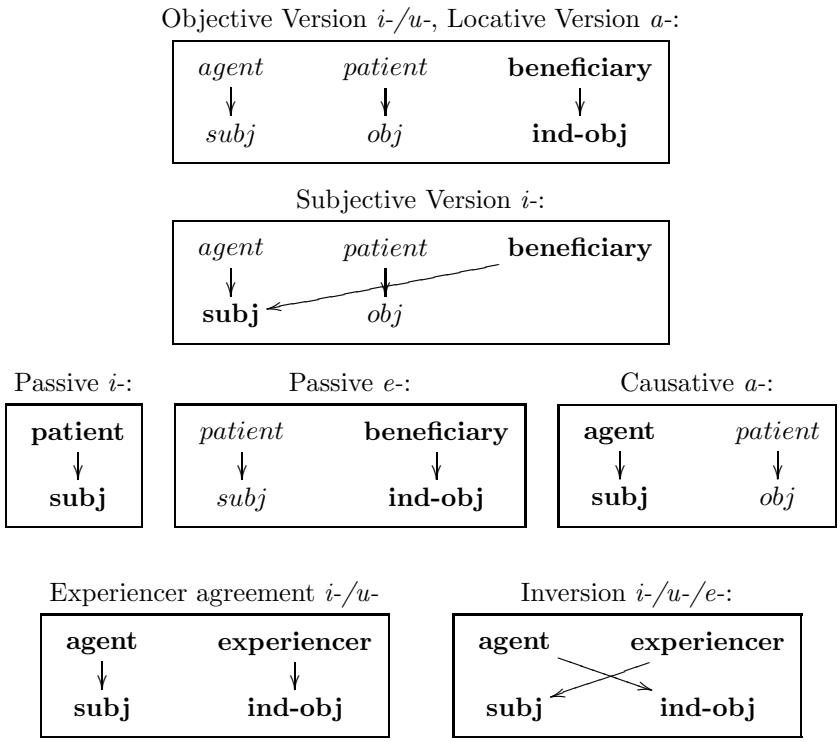


Figure 4.7: Argument-linking representation of version constructions

On the other hand, the pre-stem slot is not the only part of a verb form involved in expressing valency changes or voice. As mentioned in section 4.6, thematic suffixes (especially *-eb*), and the suffix *-in* also participate in the formation of causatives. Some passive and impersonal verbs, on the other hand, are formed by adding the suffix *d-* instead of a version vowel. In short, the valency-changing function is not uniquely associated with the pre-stem slot—there are exceptions on both formal and functional sides of the relationship.

It should also be emphasized that version is not simply part of a valence alternation that derives indirect objects from obliques. The version constructions themselves impose semantic and pragmatic constraints that should not be expected from a lexical rule-like operation. Verbs with version do not always have a corresponding versionless verb (at least synchronically), and the precise meaning of a verb with version is not always predictable based on a versionless variant.

Michaelis and Ruppenhofer (2001) come to a similar conclusion in the examination of the German applicative with the prefix *be-*. At first glance, the German *be*-pattern seems to be a simple argument promotion device parallel to the locative alternation in English, e.g. *schmieren* 'to smear X on Y' ∼ *beschmieren* 'to smear Y with X'. This pattern accepts both transitive verbs of transfer and bivalent intransitive verbs of locomotion, e.g. *wandern* 'wander', in each case promoting what would otherwise be an oblique to a core argument (usually direct object). Also, as the English locative alternation, the German applicative is usually described as applying to a limited semantic class of verbs. However, Michaelis and Ruppenhofer's extensive examination of corpus data involving this pattern reveals that the applicative pattern has its own unique properties that cannot be derived via lexical rules from non-applicative verb entries. First, in some instances the applicative pattern assigns grammatical relations that would not be expected from the input verbs. So in (154), the 'input' lexical entry 'cook' does not license a goal argument, yet the prefixed verb assigns core argument status to a semantic goal or beneficiary.

(154) Aber wenn ich es mir recht überlege habe ich mich von meinen Kollegen [...] auch mal eifrig Kaffe **bekochen** lassen [...]
'But, if I remember correctly, at times I also let myself be *be*-cooked with coffee by my colleagues.' (Michaelis and Ruppenhofer 2001:3)

Second, the corpus data also suggests that the applicative pattern is less selective

about what verbs serve as input, questioning the assumption of constrained productivity. Third, the pattern seems to accept even nouns as input, as in (155).

(155) As mag ja lustig sein, zwei hartgekochte Eier wie Clownsköpfe mit angekeimten Sojabohnen zu **behaaren** und sie auf Gurkenscheiben zu stellen, ihnen mit zwei Tomatenstreifen Münder zu verpassen und Auglein aus Sojasprossen einzudrücken. 'Ok, it might be funny to *be*-hair two hard-boiled eggs like clown's heads with germinating soy beans, to stand them up on cucumber slices, to give them mouths from tomato strips, and to impress soy shoots on them as little eyes.' (Michaelis and Ruppenhofer 2001:5)

Instead of looking for the right way to constrain the potential inputs to the applicative, Michaelis and Ruppenhofer suggest looking at the semantic and pragmatic effects of the applicative independently of its inputs. Here they find much clearer regularities. The meanings of the applicative constructions cluster around a central prototype with a sense of 'covering' where a THEME physically covers a LOCATION. Conventional metaphors like DISCOURSE IS TRAVEL OVER AN AREA, PERCEIVING IS COVERING OBJECTS WITH ONE'S GAZE, THE CONDUIT METAPHOR, EFFECTS ARE TRANSFERRED OBJECTS, described in (Reddy 1979, Lakoff and Johnson 1980, Sweetser 1990, Goldberg 1995) give rise to semantic extensions of the construction which keep the argument structure but change the semantic effect. Based on historical evidence, Michaelis and Ruppenhofer (2001) propose that the applicative pattern started from a single prototype and expanded using metaphorical links to express more general 'transfer, iterated action, intensification of the action or state denoted by the verb, and effects achieved by means of an action.' The resulting analysis is a set of applicative constructions which impose fairly specific constraints on their semantics and are more permissive with respect to input verbs / nouns, provided the inputs can be made compatible with the overall meaning.

This analysis of the German applicative is in some sense quite similar to the analysis proposed here for Georgian version. Although less is known about the semantic origins of version vowels, and there is more syntactic variation in the synchronic constructions, it seems clear that the resulting version constructions must have properties of their own not purely derivable from the versionless inputs. Evidence from the acquisition of version markers supports this hypothesis: according to Imedadze and Tuite (1992), children use version markers fairly consistently and productively, overextending them to verbs with which they

are normally restricted or not used. Thus, productivity in the version constructions seems to come from the constructions themselves and not from any inputs derived by lexical rules. These patterns strengthen the evidence for a construction-based, rather than rule-based, analysis.

The next chapter examines several instances in which pre-radical vowels are used to mark purely morphological properties, with no apparent connection to valence or voice. However, plausible patterns of historical development link many of the morphological uses to the core version uses, and such connections are examined in detail in the next chapter. Aronson (1982) examines the cases traditionally identified as examples version and fails to find conclusive evidence that it is a separate grammatical category in Georgian, as claimed by Shanidze and others after him. Aronson's conclusions are based mainly on the lack of direct paradigmatic contrasts between subjective version and reflexives, and objective version and other markings of indirect objects. The conclusions dram from this chapter depart from Aronson's in a couple of ways. On the one hand, the contrasts between objective-version and oblique phrases in Section 4.2.4 suggest that the semantic differences are psychologically real. On the other hand, both objective and subjective version are integrated into the morphosyntax of Georgian, and are in some sense parasitic on processes of voice alternations and object-marking. However, it is very hard to say what it means for something to be a "separate grammatical category," since all of the morphosyntactic processes in Georgian are so closely tied together. The next chapter reinforces the difficulty of establishing such a separation.

Chapter 5

Pre-radical Vowels and the Family of Constructions

5.1 Introduction

This chapter considers grammaticalized uses of pre-radical vowels. As a reminder, pre-radical vowels occupy the slot immediately preceding the verb root, and possible vowels are *i*-, *e*-, *a*-, and *u*-. Chapter 4 discussed semantically-motivated uses of the pre-radical vowels, such as the marking of participant affectedness, reflexivity, some types of passivization, causativization, and experiencer agreement. In certain lexical and syntactic contexts, these vowels are not optional, not semantically motivated, and serve purely grammatical, morphological purposes. The contexts include the formation of future and aorist screeves of some verbs, inversion, and the marking of lexical class often analyzed as 'neutral version.' While some of these contexts trace their origins to semantically motivated version constructions described in the previous chapter, synchronically a common distribution may be the only thing they share.

The descriptions in this chapter capture the regularities in such contexts and present the necessary contextual factors needed for constructional representations. Historical and synchronic connections are described in detail, and I suggest that the two types of connections do not always coincide. Further, I argue that commonality of distribution (form) need not imply commonality of function (meaning), although the morphological pre-radical vowels are not merely homophonous with the core version vowels. The representations in

this chapter provide a possible way of representing relationships between constructions with common form and common meaning without assuming that the two types of commonality must go hand-in-hand.

As before, the behavior or pre-radical vowels is stable within a given combination of series and conjugation class.

5.2 Pre-Radical Vowels as Stem and Screeve Formants

The pre-radical vowels *i-* and *e-* in some constructions serve as markers of the screeve, or TAM paradigm. The vowel *i-* shows up in future and aorist forms of Conjugation 3 (unergative) verbs, and in the future forms is sometimes the only morphological feature differentiating them from the present forms. The vowel *e-* can mark future in some Conjugation 2 verbs and some Conjugation 4 verbs; both these vowels serve as stem formants similar to those described for Latin and Hebrew by Aronoff (1994). This section explores these uses in detail and suggests that the screeve-marking function is not synchronically related to the core participant affectedness uses described in Chapter 4.

5.2.1 Conjugation 3 (Unergative) Verbs

In Conjugation 3 (Unergative) Verbs, the future and aorist stem is marked by the addition of the pre-radical vowel *i-*. For some verbs, this vowel is the only feature that distinguishes the present for from the future form (156). By contrast, Conjugation 1 verbs form their future using preverbs, as described in section 3.3.2. Since the future and aorist stems are interdependent, one cannot say that the vowel *i-* 'means' future tense. Rather, it is an inherent part of the future and aorist-stem formations for Conjugation 3 verbs. Holisky (1981)a identifies such verbs as atelic, or *activities* in the Vendler's (Vendler 1967) classification.

(156) a. *pren / pren-s*
fly / fly-SCR
'You are flying / S(he) is flying (PRESENT)'

b. ***i**-pren / **i**-pren-s*
prv-fly / **prv**-fly-SCR
'You will fly / (S)he will fly (FUTURE)'

c. ***i***-*prin-e* / ***i***-*prin-a*
prv-fly-SCR / **prv**-fly-SCR
'You flew / (S)he flew (AORIST)'

For many Conjugation 3 verbs, the *i-* is accompanied by a specific thematic suffix in the future, which is different from the thematic suffix that the verb takes in the Present. Most often, this is the suffix *-eb*, as in (157).

(157) a. *cek'v-**av***
dance-**ts**
'You dance'

b. ***i***-*cek'v-**eb***
prv-dance-**ts**
'You will dance'

Another possible future construction involves the suffix *-i* along with the pre-radical vowel (158).

(158) a. *sc'avl-ob*
learn-TS
'You are learning'

b. ***i***-*sc'avl-**i***
prv-learn-**ts**
'You will be learning'

In any case, the pre-radical vowel *i-* serves as the invariant part in the formation of the future/aorist stem in Conjugation 3 verbs. It has the same distribution as the subjective-version vowel *i-*, but none of the same functions. In many Conjugation 3 verbs, the version vowel *i-* can be used in its version meaning as well (159). As mentioned in Chapter 4, only objective (and not subjective) version is possible with Conjugation 3 verbs. This is not a phonological effect of disallowing two identical vowels next to each other, as evidenced by words like *iisperi* 'violet-colored', *k'aliisxela* 'the size of a grasshopper'. Rather, there appears to be a slot constraint that only one pre-radical vowel is allowed in the verb template.

(159) a. *shen m-**i***-*cek'v-av*
you 1SGOBJ-**prv**-dance-TS
'You dance for me (OBJECTIVE VERSION).' (PRESENT)

b. *shen* ***i****-cek'v-eb* *chem-tvis*
you **prv**-dance-TS 1SG-for
'You will dance for me.' (FUTURE)

c. *shen (me) m-****i****-cek'v-eb*
you (I) 1SGOBJ-**prv**-dance-TS
'You will dance for me (OBJECTIVE VERSION).' (FUTURE)

Some scholars (e.g. Tschenkéli 1958) suggest that historically, the future forms of unergative verbs were derived form subjective-version forms of transitive verbs. When Conjugation 3 verbs were first developed, they did not have future forms, and when those became needed they were 'borrowed' from subjective-version Conjugation 1 forms. However, Nozadze (1974) and Harris (1981) have argued against this development. Nozadze suggests that instead, the *i-* stems are quite old and were previously used in all screeve forms of unergatives. Such an analysis could help explain the affinity of unergative verbs and middle voice (as discussed in Section 5.7.1), but does not explain the synchronic function of the pre-radical vowels in Conjugation 3 forms[1]

5.2.2 Conjugation 2 (Unaccusative) Verbs

The pre-radical vowel *e-* can mark future in some Conjugation 2 (unaccusative) verbs. This use creates forms which are ambiguous between future forms and present-tense forms with an indirect object, as described in section 4.5. The two uses can sometimes be disambiguated by the argument structure, but given Georgian's propensity for pro-drop, the ambiguity remains quite real.

(160) a. *i-xat'-eba*
PRV-paint-TS
'It is painted.' (PRESENT)

b. *e-xat'-eba*
PRV-paint-TS
'It will be painted.' (FUTURE)

c. *e-xat'-eba* *mas*
PRV-paint-TS 3SG.DAT
'It is painted for/in front of him/her.' (PRESENT)

[1]See Harris (1981:234) for some additional objections to Tschenkéli's borrowing analysis.

Constructionally, the use of *e-* in future constructions can be represented as in 161.

(161)
$$\begin{bmatrix} \textit{cxn-Future-Conj2} \\ \text{SEM} \quad \begin{bmatrix} \text{TENSE} & \text{FUTURE} \end{bmatrix} \\ \text{SYN} \quad [] \\ \text{MORPH} \quad \begin{bmatrix} \ldots + \text{e} + \text{ROOT} \ldots \end{bmatrix} \end{bmatrix}$$

5.3 Inversion

The term "inversion" refers to a construction in which the semantic agent is marked by dative case and is cross-referenced on the verb by *m*-set markers, and the semantic patient appears in the nominative case and is cross-referenced on the verb by *v*-set markers. This phenomenon occurs ccur in series III (perfect and pluperfect screeves) of active verbs (i.e. Conjugation 1 and 3) and in all forms of Conjugation 4 ('indirect') verbs, in Harris' classification. Semantically, the inversion construction de-emphasizes the agentivity of the subject. Inversion in series III coincides with an evidential interpretation in affirmative perfect and pluperfect. Indirect verbs do not have evidential semantics, but tend to include patient-like verbs, e.g. verbs of experience, perception, and possession. As argued in Section 3.2.6, inversion represents a different mapping between thematic roles and grammatical arguments. In addition to the syntactic differences with non-inverted constructions, inverted forms are also characterized by the presence of pre-radical vowels.

5.3.1 Inversion in Active Verbs

Inversion applies to series III of Conjugation 1 and Conjugation 3 verbs (transitive and unergative). Instead of normal version markers, the pre-stem slot in these verbs has obligatory markers *i/u-* or *e-*, which indicate screeve (TAM) in series III. In the perfect, forms with a 1st and 2nd person initial subject have the pre-radical vowel *i-* (162a); forms with a 3rd person initial subject have the pre-radical vowel *u-* (162b). A beneficiary of the action can only be expressed in a postpositional phrase.

(162) a. *(turme) saxl-i a-m-**i**-sheneb-ia.*
(apparently) house.NOM PV-1SGOBJ-**prv**-build-PERF
'I have (apparently) built a house.'

b. *mas saxl-i a-**u**-sheneb-ia.*
3SG.DAT house.NOM PV-**prv**-build-PERF
'(S)he has built a house.'

In the pluperfect screeves, all forms have the pre-radical vowel *e-* (163).

(163) a. *saxl-i (amxanag-is-tvis) unda a-m-**e**-sheneb-ina.*
house.NOM (friend-GEN-for) must PV-1SGOBJ-**prv**-build-PLUPERF
'I was supposed to have built a house.'

b. *saxl-i (amxanag-is-tvis) unda a-**e**-sheneb-ina.*
house.NOM (friend-GEN-for) must PV-**prv**-build-PLUPERF
'(S)he was supposed to have built a house.'

Subjective or objective versions are not possible in inversion (164).

(164) * *saxl-i amxanag-s unda a-m-**i/u**-sheneb-ina.*
house.NOM friend-DAT must PV-1SGOBJ-**prv**-build-PLUPERF
'I was supposed to have built a house for a friend.'

Another group of verbs that exhibit inversion are the so-called passives of state (Cherchi 1997, Aronson 1990). This is an unproductive and marginal group that express "a particular state of the Nominative subject" (Hewitt 1995:393b); they are typically grouped with indirect verbs in Conjugation 4. Passives of state usually correspond to active, non-stative verbs and are characterized by the ending *-ia* in the Present (165a). When such verbs have an indirect-object beneficiary or possessive, they also add a pre-radical version *i-/u-* (165b), depending on the person of the indirect object, or sometimes the prefixes *s-h-*.

(165) (Aronson 1990)

a. *magida-ze axali supra pen-ia.*
table-on new tablecloth.NOM spread-SCR
'A new tablecloth is spread on the table.'

b. *magida-ze axali supra gv-i-pen-ia.*
table-on new tablecloth.NOM 1PLOBJ-PRV-spread-SCR
'A new tablecloth is spread on *our* table.'

Passives of state are aspectually atelic and therefore do not have a distinct Aorist tense (given that Aorists are normally perfective). In the past tense, these verbs are marked by the *e-* pre-radical vowel and the Aorist endings (*-a* or *-o*) of the corresponding active verbs (166).

(166) a. *magida-ze axali supra* *e-pin-a.*
table-on new tablecloth.NOM PRV-spread-SCR
'A new tablecloth was spread on the table.'

b. *magida-ze axali supra* *gv-e-pin-a.*
table-on new tablecloth.NOM 1PLOBJ-PRV-spread-SCR
'A new tablecloth was spread on *our* table.' (Aronson 1990)

As a result, Present and Past forms of passives of state are often homophonous with corresponding Perfect and Pluperfect forms of active verbs, respectively. Passives of state are older in the development of the Georgian verbal system than series III Perfect and Pluperfect tenses (Harris 1985), and so it is possible that the use of pre-radical vowels, and inversion in general, was applied by analogy from passives of state to series III tenses.[2]

In inversion, the agent/actor appears to have some the formal properties of an indirect object. The use of pre-radical vowels may have arisen through the types of verbs described in section 4.7, where version markers *i/u-* became grammaticalized as marking indirect object agreement without regard for "versional" meaning of affectedness. However, the agreement function of the pre-radical vowels does not explain why *i/u-* is used in the perfect and *e-* is used in the pluperfect. A synchronic description must describe the use of pre-radical vowels in inversion separately, and the constructional analysis presented here provides a syntactic description of the synchronic system.

A constructional description would simply state that the pre-radical vowels are required parts in the formation of the perfect and pluperfect screeves, as in (167) and (168), along with the inverted syntax and other screeve-formation morphology.

(167)
$$\begin{bmatrix} \textit{cxn-Perf-Conj1} & \\ \text{SEM} & \begin{bmatrix} \text{TENSE} & \text{PERFECT} \\ \text{MOOD} & \text{EVIDENTIAL} \end{bmatrix} \\ \text{SYN} & [] \\ \text{MORPH} & \begin{bmatrix} \ldots + \text{i/u} + \text{ROOT} \ldots \end{bmatrix} \end{bmatrix}$$

[2] I would like to thank Kevin Tuite for bringing my attention to point.

(168)
$$\begin{bmatrix} \textit{cxn-PluPerf-Conj1} & \\ \text{SEM} & \begin{bmatrix} \text{TENSE} & \text{PLUPERFECT} \\ \text{MOOD} & \text{EVIDENTIAL} \end{bmatrix} \\ \text{SYN} & [] \\ \text{MORPH} & \begin{bmatrix} \ldots + \text{e} + \text{ROOT} \ldots \end{bmatrix} \end{bmatrix}$$

5.3.2 Inversion in Conjugation 4 (Indirect) Verbs

Conjugation 4 ("Indirect") verbs also have dative-marked agents and, if transitive, nominative marked patients. Some indirect verbs can have an obligatory marker *i/u-* or *e-* in series I and II (169). Their series III forms are either participial and thus do not contain pre-radical vowels, or contain markers from the *s/h-* set (170).

(169) a. *m-**i**-q'vars / **u**-q'vars*
1SGOBJ-**prv**-love / **prv**-love
'I love it / (S)he loves it.'

b. *m-**e**-mγer-eba*
1SGOBJ-**prv**-sing-PASS
'I feel like singing.'

(170) *m-q'var-eb-ia / h-q'var-eb-ia*
1SGOBJ-s/h-love-TS-SCR / 3SGOBJ-s/h-love-TS-SCR
'I have loved it / (S)he has loved it.'

Other indirect verbs use object markers and *s/h-* (171), although the future tends to be marked by *e-* for all indirect verbs (172).

(171) *mo-**m**-c'ons / mo-**s**-c'ons*
PV-1SG-like / PV-3SG-like
'I like it / (S)he likes it.'

(172) a. *mo-**e**-c'oneba*
PV-PRV-like
'He will like it.'

b. ***e**-q'vareba*
PV-PRV-love
'She will love it.'

The contrast between *i/u-* and *s/h-* is reminiscent of the contrast between these markers in grammaticalized patient marking, which is fairly lexically dependent (as noted by Aronson 1982). In the case of indirect verbs, the motivation for using one construction over another seems to be even more lexicalized, and it does not seem possible in synchronic terms to provide a semantic distinction between them. A constructional representation, again, would have to specify that the pre-radical vowels (or consonants) are a required part of the verb formations as in (173) and (174).

(173)
$$\begin{bmatrix} \textit{cxn-Conj4-type1-present} \\ \text{SEM} \quad \begin{bmatrix} \text{TENSE} \quad \text{PRESENT} \end{bmatrix} \\ \text{SYN} \quad [] \\ \text{MORPH} \quad \begin{bmatrix} \ldots + \text{i/u} + \text{ROOT} \ldots \end{bmatrix} \end{bmatrix}$$

(174)
$$\begin{bmatrix} \textit{cxn-Conj4-type2-present} \\ \text{SEM} \quad \begin{bmatrix} \text{TENSE} \quad \text{PRESENT} \end{bmatrix} \\ \text{SYN} \quad [] \\ \text{MORPH} \quad \begin{bmatrix} \ldots + \text{s/h} + \text{ROOT} \ldots \end{bmatrix} \end{bmatrix}$$

5.4 'Neutral' version *a-*

As discussed in Chapter 4, the pre-radical version *a-* and thematic suffix *-eb* (sometimes *-ob*) are used to form morphological causatives. The same pattern marks a lexical class of verbs, many of which are derived from noun and adjective roots (175). This class contains around 2700 verbs and is quite productive (Melikishvili 2001). Historically, the use of *a-* in these cases may signal a verbalizing function, but many of the verbs in this class no longer share a synchronic connection with the stem from which they were derived. Moreover, many of the verbs in this class are not denominal or deadjectival (176). In the questionnaire described in Chapter 3, the majority of respondents used a version vowel when producing verbs with the thematic suffix *-eb*, even though the question did not prompt them to do so, and the nonce verbs they were given were certainly not denominal.

(175) Verbs derived from nouns and adjectives

a-rig-eb 'arrange'	*rigi* 'order'
a-γmert-eb 'deify'	*γmerti* 'god'
a-γarib-eb 'impoverish'	*γarib* 'poor'
a-c'itl-eb 'redden'	*c'iteli* 'red'
a-partov-eb 'broaden'	*parto* 'broad'

(Aronson 1990:72), (Hewitt 1995b:219)

(176) Verbs not derived from nouns or adjectives

a-shen-eb 'build'	*shen (not a noun)
a-k'et-eb 'do'	*k'et (not a noun)
a-shr-ob 'dry'	*shr (not a noun)

In this class of verbs, the version vowel *a-* often occurs in complementary distribution with other version vowels. Most verbs in this category belong to Conjugation 1, as in (177a). When the pre-stem slot is occupied by version vowel *a-*, it cannot express version-like meanings, so that beneficiaries of actions have to be expressed in postpositional phrases (177b). However *a-* can be replaced by other version vowels to form objective or subjective version, as in (177c). This, again, is not due to phonological reduction of multiple vowels, as evidenced by words like *aasheneb* 'you will build' (*a-* is a preverb), *ia* 'violet', *ai* 'here is'.

(177) a. *saxl-s* ***a**-shen-**eb***
house-DAT **prv**-build-**ts**
'You build a house (beneficiary not specified).'

b. *saxl-s* ***a**-shen-eb* *mis-tvis* */ *mas*
house-DAT **prv**-build-TS he.GEN-for / *he.DAT
'You build a house for him.'

c. *saxl-s* ***u**-shen-eb* **mis-tvis* */ *mas*
house-DAT **prv**-build-TS *he.GEN-for / he.DAT
'You build a house for him.'

The verbs with pre-radical *a-* includes transitive verbs like *a-mtknar-eb* 'you yawn', and even indirect verbs like *m-a-vicq'deba* 'I forget'.

Because it stands in contrast with other possible version vowels, *a-* has been called 'neutral version' in the Georgian linguistic tradition (Boeder 1968, Hewitt 1995b). However, its meaning and function are not related to the 'participant affectedness' meanings of the other versionizers, and it is not required on all verbs.

5.5 The family of constructions

As demonstrated in the previous sections, pre-radical vowels participate in several distinct types of constructions with distinct functions. Some of the uses are in paradigmatic opposition with each other (e.g. subjective vs. objective vs. locative version), while others are synchronically related (e.g. objective version and grammaticalized patient marking). Figure 5.1 represents these connections. The solid lines represent paradigmatic oppositions; arrows represent synchronic semantic connections. Finally, all of the version constructions are related via similarity of form (i.e. the use of pre-radical vowels). Such connections are not paradigmatic because the morphosyntactic properties defining the different constructions are not sufficiently similar (e.g., the use of *i-* as objective version vs. future marker). However, distributional similarities still provide a connection between these diverse constructions in which pre-radical vowels participate.

In addition, some of the version uses are related historically, as we know from historical developments; other uses may plausibly be related historically, although there is no direct evidence for these relationships. Figure 5.2 represents the confirmed (solid arrows) and plausible (dashed arrows) historical connections between version uses. As the diagrams demonstrate, historical connections do not always equal synchronic connections; in several cases, confirmed historical connections no longer carry synchronic meaning.

The functional classes of pre-radical vowels are summarized in Section 5.6.

5.6 Summary of functional classes of version

1. Active Constructions: objective version (*i/u-*) vs. subjective version (*i-*) vs. locative version (*a-*)).
    - Applies to: Transitive and unergative verbs in series I and II, non-prefixal Conjugation 2.
    - Semantic function: Mark the relationship between the action and an affected party (usually indirect object).
    - Syntactic function: May add an indirect object to the verb's argument structure where it is not otherwise licensed.

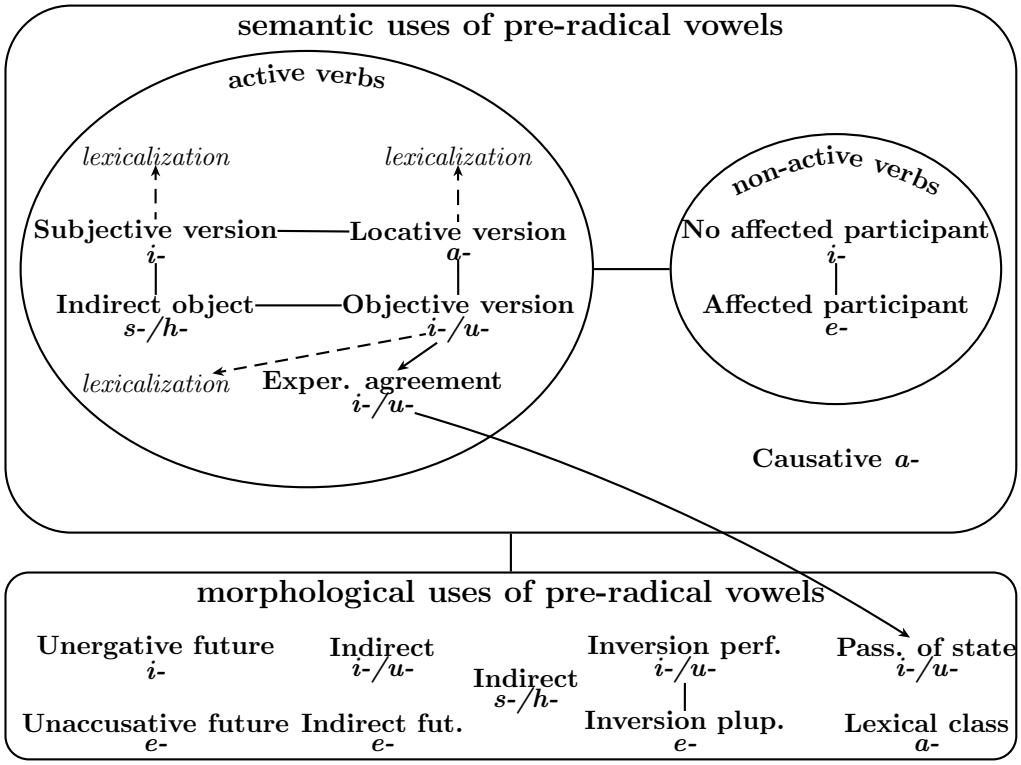


Figure 5.1: Uses of pre-radical vowels: Synchronic groups and connections

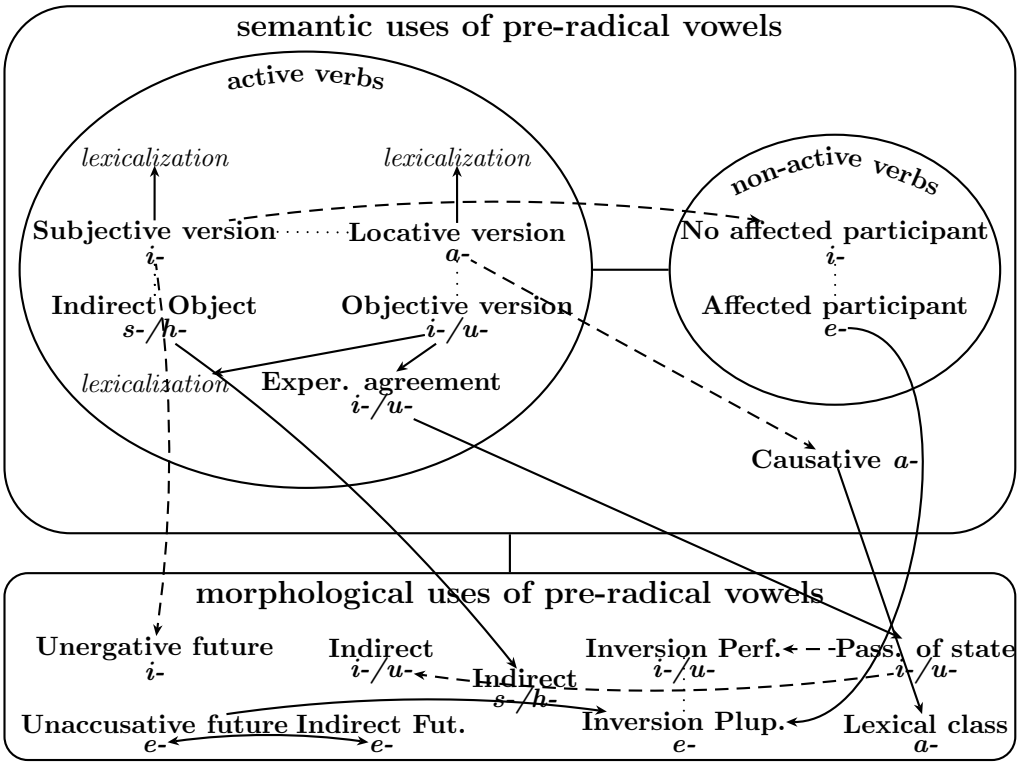


Figure 5.2: Uses of pre-radical vowels: Confirmed and Putative Historical Connections

- Degree of productivity: Objective version is very productive, subjective version is somewhat less productive, locative version is the least productive.
- Lexicalization: All three versions have become lexicalized to some extent, especially the subjective version. Some verbs with originally reflexive meanings now require the subjective versionizer at all times.

2. Experiencer agreement: *i-/u-* or *s-/h-*

   - Applies to: Conjugation 1 verbs that license an indirect object.
   - Semantic function: mark indirect object agreement.
   - Syntactic function: mark indirect object agreement.
   - Degree of productivity: used with many ditransitive verbs.
   - Lexicalization: some verbs prefer *s-/h-*, others prefer *i-/u-*, unpredictably.
   - Probably derives from Class 1.

3. Unaccusative / Passive Constructions: no indirect object (*i-*) vs. with indirect object

(*e-*).

- Applies to: unaccusative verbs in series I and II.
- Semantic function: Mark the presence or absence of an indirect object.
- Syntactic function: *i-* indicates the absence of an affected participant; *e-* indicates the presence of an affected participant.
- Degree of productivity: Very common way of forming Conjugation 2 verbs.
- Lexicalization: For some verbs (so-called bipersonal passives), an indirect object is required, and *e-* is the only possible versionizer.

4. Passives of state, present (*i-, u-, s-, h-*; *e-* in the past).
    - Applies to: passives of state.
    - Semantic function: express a state.
    - Syntactic function: subject in the nominative, indirect object / experiencer in the dative.
    - Degree of productivity: Very limited.
    - Lexicalization: this is a closed class.
    - Use of versionizers may be derived from Class 2.
5. Indirect verbs
    - Applies to: Conjugation 4 verbs.
    - Semantic function: the main actant is an experiencer rather than agent.
    - Syntactic function: experiencers mapped to indirect objects, stimuli mapped to subjects.
    - Degree of productivity: closed class.
    - Morphological form probably derives from Class 4.
6. Inverse Series III: perfect tense (*i/u-*) vs. pluperfect tense (*e-*).
    - Applies to: Transitive and unergative verbs in series III; some indirect verbs in all series.
    - Semantic function: Mark the loss of control by the agent, as well as tense (perfect vs. pluperfect) for active verbs.
    - Syntactic function: Indicate that the agent is mapped to indirect object.
    - Degree of productivity: Productive for all transitive verbs; the indirect verb class is not productive.

- Lexicalization: Versionizers are obligatory, so there is no lexicalization.
- Versionizers derived from Class 2.

7. Future and Aorist tenses of unergative verbs: *i-*.

   - Applies to: unergative verbs, series I (Future sub-series) and II.
   - Semantic function: Mark tense.
   - Syntactic function: None.
   - Degree of productivity: All unergative verbs use this pattern.
   - Lexicalization: Lexicalized with some Conjugation 1 verbs.
   - Versionizers may be derived from Class 1.

8. Future and Aorist tenses of unaccusative, indirect verbs: *e-*.

   - Applies to: unaccusative and indirect verbs, Future and Aorist.
   - Semantic function: Mark tense.
   - Syntactic function: None.
   - Degree of productivity: All unaccusative and many indirect verbs use this pattern.
   - Versionizers may be derived from Class 2.

9. Causatives: *a-*.

   - Applies to: transitive and unergative verbs.
   - Semantic function: Add a causer.
   - Syntactic function: Add another argument (the agent) and demote the old arguments. The versionizer is *a-* and a causative suffix *(-in)-eb*.
   - Degree of productivity: Fairly productive.
   - Lexicalization: In some verbs, required even while not accompanied by the suffix *-eb*.
   - May be related to locative version from Class 1.

10. Morphological *a-*.

    - Applies to: many transitive verbs.
    - Semantic function: None.
    - Syntactic function: None.

- Degree of productivity: Very productive with verbs derived from nouns and adjectives.
- Lexicalization: In some verbs, required even while not accompanied by the suffix *-eb*.
- Versionizer possibly derived from Class 5.

5.7 Concluding Discussion

5.7.1 Version and Middle Voice

The diverse functions of pre-radical vowels in subjective version in Conjugation 1 and passive/impersonal verbs in Conjugation 2 resemble the range of functions often filled by middle voice, as described by Klaiman (1991) and Kemmer (1994). The similarity between subjective version and middle voice has also been pointed out by Shanidze (1953:362-363) and Schmidt (1965). Tuite (2002, 2003) explores this relationship further. The parallels between subjective version and middle voice are seen most clearly in Table 5.1. However, Georgian objective version finds no parallel category in Indo-European.

GREEK MIDDLE DIATHESIS	GEORGIAN SUBJECTIVE VERSION
*louo-**mai*** "I wash myself"	*da-v-**i**-ban* "I will wash my (hands, self)"
*hetoimazo-**mai*** "I prepare (for) myself"	*mo-v-**i**-mzad-eb* "I will prepare myself"
*hapto-**mai*** "I grasp, fasten myself to"	*da-v-**i**-b-am (tavs)* "I will bind myself"
*kikhra-**mai*** "I borrow"	*v-**i**-sesx-eb* "I will borrow"
*ôneo-**mai*** "I buy"	*v-**i**-q'id-i* "I will buy"
*orkheo-**mai*** "I dance"	*v-**i**-cekv-eb* "I will dance"

(Tuite 2003)

Table 5.1: Greek middle diathesis and Georgian subjective version

In addition, Tuite examines the Georgian Conjugation 3 (unergative / medioactive) verbs, which have been classified by Holisky (1981a) as in Table 5.2. Verbs with similar semantics typically belong to the middle voice, or *media tantum*, group in Greek. Present-tense conjugations of these verbs do not reveal any morphological affinity with other 'middle voice' formations like subjective version of passives. However, the future and aorist tenses of these verbs are formed by adding the pre-radical vowel *i-*. As men-

tioned earlier, the traditional analysis (cf. Tschenkéli 1958) suggests that the future forms 'borrowed' morphology from the subjective-version forms of transitive verbs, although this seems unmotivated. According to Nozadze (1974:43-44) and Tuite (1997), there may indeed be a historical relationship between these two groups of verbs, but possibly in a different direction. Medioactive verbs in Svan and Mingrelian use pre-radical vowels in all tense forms, and Nozadze suggests that Old Georgian medioactives behaved the same way. Thus, the presence of pre-radical vowels in medioactives may have been related to middle voice functions, and could possibly have been extended to subjective-version and passive verbs.

(i)	*celk-ob-s* "behaves naughtily" (denominal: "behave like X")
(ii)	*sadil-ob-s* "eats noontime meal" (denom.: "partake of / participate in / experience X)
(iii)	*gor-a-ob-s* "rolls around" (deverbal; unidirected motion or action)
(iv-a)	*sisin-eb-s* "hisses (of a snake)" (sounds)
(iv-b)	*q'iv-is* "crows, screeches" (animal cries)
(v)	*gizgiz-eb-s* "flickers" (qualities of emitted or reflected light)
(vi)	*k'ank'al-eb-s* "trembles" (motion in place)
(vii)	*gor-av-s* "rolls" (manner of motion)
(viii)	*kux-s* "thunders" (natural, weather phenomena)
(ix)	Residual (older, irregular verbs, e.g. *i-cin-i-s* "smiles")

(Tuite 2003, based on Holisky 1981a)

Table 5.2: Subgroups of Georgian medioactive verb classes

Following Shanidze (1953:305) (as well as Jorbenadze 1983 and Sarjveladze 1987), Tuite (2002) identifies some of the Conjugation 2 verbs as deponents, or "passive in form and active in meaning." These deponent verbs, functionally quite similar to Greek middles, can be divided into several semantic classes as in (Tuite 2002). One of their primary semantic characteristics, just as with subjective-version and medioactive verbs, is to profile or draw attention to the sole argument subject / undergoer of the action, which is the typical effect of middle voice in Greek or Latin. Thus, Tuite suggests that the Greek category of middle voice finds significant correlates in Georgian, albeit in three different types of formations (subjective voice, passives / deponents, and medioactives). In some cases synchronically and in some cases historically, all of these formations are affiliated

with the use of pre-radical vowels. However, there are also significant differences between the Greek and Georgian situations. The Georgian parallels to middle voice are much more heterogenous than the Greek ones appear to be, and more lexicalized. It is not clear whether there was ever a unified notion of middle voice in Georgian, or if 'middle voice-marking' pre-radical vowels were extended from a relatively small group of medioactive verbs to cover semantically related ground in other verbs. Moreover, the Georgian middle-voice formations cover only a portion of the semantic classes described for Greek middle voice (Kemmer 1994, Tuite 2003), and it's not clear whether there is a synchronic category of 'middle voice' in Georgian.

5.7.2 Distributional similarity does not imply functional similarity

Pre-radical vowels occupy the same pre-stem slot on the verb, but they serve radically different functions in different contexts. These functions are motivated by the larger morphosyntactic constructions, but the morphosyntactic constructions do not uniquely determine the functions of pre-radical vowels.

These distributional facts present problems for approaches that assume that meaning is built from the bottom up, from the smallest pieces to the larger pieces. In such an approach, it is natural to assume that the same form should have the same meaning in all contexts (homonymy excluded), and the same distribution class should have the same function. Under such an approach, the vowel *a-* in verbs like *a-sheneb* 'you build' stands in opposition to the subjective-pre-radical vowel *i-* in *i-sheneb* 'you build for yourself' and should therefore have a contrasting but related function. As a result, *a-* has been analyzed as indicating 'neutral version', producing the puzzling fact that only some verbs require it, and the category means nothing.

Similarly, the historical connections between some uses of pre-radical vowels have led to the analysis of Conjugation 3 future forms as somehow related to reflexives (subjective-version forms). This analysis, fails to explain why the two sets of forms are so different in terms of their syntax and semantics, however.

There are two parts to fixing this problem. First, one should separate the formal patterns from the meaning patterns and realize that they may have separate principles of organization. The match between form and meaning may occur at the smallest level (that of a morpheme), but it is just as likely, if not more so, to occur at the level of the word or

phrase. If the meaning is determined at the level of a word or phrase, then individual pieces of form and meaning need not be in a one-to-one relationship to each other. As argued in the Chapter 3, the optimal constructional level for Georgian is that of the series. Given the series and verb conjugation type, the function of pre-radical vowels is quite clear.

Second, one should keep historical facts separate from synchronic facts. Most of the functional classes of pre-radical vowels are, indeed, historically related. However, many of these historical connections are no longer manifest in the synchronic uses of pre-radical vowels, and are unlikely to be known to ordinary speakers. Thus it makes more sense to talk about specific constructions in which pre-radical vowels are used than to assign unique meaning to each pre-radical vowel. On the other hand, historical facts may be essential for *explaining* peculiar synchronic patterns, as is argued extensively by Juliette Blevins (2004) and Harris (2004).

5.7.3 Regarding Compositionality

Pre-radical vowels in Georgian demonstrate the shortcomings of a naive view of compositionality. Under such a view, meaning is attributed to small parts of constructions (e.g., morphemes or words), and a few straightforward rules of composition are sufficient to derive the meaning of a whole word or phrase. The only exceptions to this principle are considered to be idiomatic.

However, the most important aspect of meaning in Georgian is *the way* in which elements combine into larger wholes. The knowledge of possible combinations of pre-radical vowels and the rest of a verbal form is essential for stating the meaning of verb forms. It may still be possible to assign specific meanings to individual morphs (e.g., *i-* as marking subjective version), but only given the formal and semantic properties of a larger construction (e.g., active transitive verbs).

A naive view of compositionality, when faced with such a conundrum, is forced to postulate homonymous morphs or diacritics. But given the historical relationships between different uses of pre-radical vowels, such an analysis misses important generalizations. On the other hand, relating constructions as a whole would allow us to note possible relationships while keeping synchronic uses sufficiently separate.

Finally, pre-radical and version vowels are best described using a family of constructions which incorporate facts about form as well as meaning. The 'meaning' part

of these constructions must involve discourse, not just syntax or literal semantics. A construction-based approach provides a natural medium in which such complex relations can be expressed, and is particularly well-suited for representing sub-regularities within a larger set of patterns.

The following chapter turns to a different extension of the Constructional Morphology framework, one involving computational modeling.

Chapter 6

Computational Modeling

6.1 Introduction

One of the frequent extensions and validations of linguistic analysis is computational modeling, and the two areas exhibit important mutual feedback relations. One the one hand, computational models underlie practical applications such as text processing, generation, and tools for language learning. On the other hand, the adaptation of a linguistic analysis for a computational model often makes the analysis more precise and can clarify otherwise unnoticed inconsistencies and mistakes in the analysis.

Computational models tend to correspond to particular linguistic theories, and the types of patterns they can represent are therefore constrained by the types of analyses proposed by the theories. Many currently widespread approaches to computational linguistics tend to incorporate the same limiting assumptions that characterize "bottom-up", or generative approaches described in Chapter 2. Thus, computational models tend to be limited in the sorts of patterns they can produce or recognize; on the other hand, some decisions in computationally-friendly linguistic theories tend to be made on the basis of technical limitations of associated computational models.

In morphological modeling, the interdependency between computation and linguistic theory often translates into computational models that consist of collections of morphemes and sets of rules for putting them together so as to generate / recognize all and only possible word forms. This chapter examines the possibilities and limitations of current approaches to computational morphology in light of the Georgian data described in Chapter 3. The issues to be addressed are the nature of a desired morphological analyzer for

Georgian, the types of patterns it must be able to address, and the fundamental limitations it runs up against given current technological possibilities.

As proof of concept, I present a prototype model of Georgian inflectional morphology in the widespread computational framework of Finite-State Transducers (FST; Beesley and Karttunen 2003), the state-of-the-art in computational modeling. The prototype model is capable of producing a large subset of the patterns described for Georgian. The model somewhat extends the possibilities of a typical finite-state transducer by combining several transducers into a sequence, thus allowing for the expression of some non-compositional patterns such as 'parasitic' or Priscianic formations (Matthews 1991), discontinuous expression of verb agreement, and the separation of lexically-determined from sub-regular from regular patterns of inflectional formation. At the same time, the model is fundamentally limited by FST's inability to express paradigmatic contrasts (except via dummy co-indexing elements that could trigger them). Earlier versions of this model are described in (Gurevich 2005a, Gurevich 2006).

The prototype computational model of Georgian has been used in the creation of an online reference on Georgian verb conjugations for language learners. Such a resource is a virtual necessity for someone learning Georgian due to a lack of comprehensive references on verb formation. Most grammars and textbooks provide only partial rules, which often stop short of predicting the formation of full paradigms of specific verbs. Recognition is as much of a problem: getting a dictionary form (usually the masdar) from a fully inflected verb form is often not a straightforward task, particularly if only one of the two forms has a preverb. From a theoretical standpoint, the online reference aims to improve second-language acquisition by increasing students' exposure to inflectional morphology.[1]

It should be kept in mind that FST models have no established 'psychological reality', and are useful largely as models of production and recognition of closed morphological models for purposes of natural language engineering. They are not designed to represent human productivity and are incapable of extending existing patterns to novel items. However, this limitation also means that changing or abandoning aspects of the FST model affects only the linguistic applications and has no bearing on psychological plausibility of any associated linguistic theories.

I begin with a review of the prevailing approaches to computational morphology,

[1]For an example of a principal parts / exemplary paradigm-based learners reference/dictionary for Estonian, see the online version of (Erelt 1999) at <http://www.eki.ee/dict/QS2006.tegemisel/>

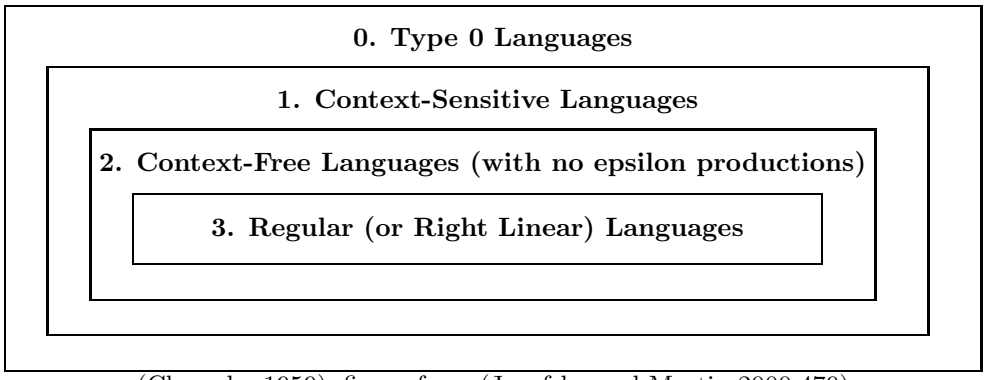
including finite-state technologies. Following a brief reminder on what makes Georgian challenging for the compositional, morpheme-based models, I describe the proposed computational model and compare it with some previous proposals for Georgian. Finally, a practical application of the computational model is described: an online Georgian verb reference for language learners.

6.2 Approaches to Computational Morphology

The basic task of computational morphologies is to be able to recognize related words. Earlier natural language processing operations relied on *stemming*, or stripping words of their morphology in order to recognize the common stems. For instance, words like *operate*, *operating*, and *operated* would be stripped down to the stem *operate* for purposes of syntactic and semantic parsing. However, stemming was shown not to improve parsing significantly, in part because the stripped morphology carries information that is lost during stemming (Manning and Schütze 1999:132). So for languages with relatively simple morphology (like English), it makes more sense to keep a list of all possible words. However, the listing strategy is obviously less plausible for morphologically complex languages. While the physical limitations such as the amount of storage available and the efficiency of search and retrieval procedures are continually being pushed further, the manual task of listing all of the forms would stump most developers of large-scale morphological models for such languages.

Thus, morphological analyzers were designed from the beginning to deal with languages other than English. From the earliest models, it was suggested that morphological alternations can be represented as *regular expressions*, or a set of statements that can be described by a *regular language* in the Chomsky language hierarchy (Chomsky 1959). Chomsky proposed a classification of formal grammars, or sets of rules describing formal languages, into four classes based on their expressiveness, or generative power. Grammars of greater generative power can generate all strings in languages of lesser generative power, as the diagram in Figure 6.1 demonstratesl Type 0 grammars are the most powerful and regular grammars are the least powerful.

The four classes of grammars correspond to constraints on how grammar rules can be written, and what computational mechanism can generate the corresponding languages, as in Figure 6.2. The distinctions and definitions of particular grammar types are discussed



(Chomsky 1959); figure from (Jurafsky and Martin 2000:479)

Figure 6.1: A Venn diagram of the languages on the Chomsky Hierarchy

TYPE	COMMON NAME	RULE SKELETON	LINGUISTIC EXAMPLE
0	Turing Equivalent	$\alpha \rightarrow \beta$, s.t. $\alpha \neq \varepsilon$	Augmented Transition Networks
1	Context Sensitive	$\alpha A \beta \rightarrow \alpha \gamma \beta$, s.t. $\gamma \neq \varepsilon$	Tree-Adjoining Grammars
2	Context Free	$A \rightarrow \gamma$	Phrase Structure Grammars
3	Regular	$A \rightarrow xB$ or $A \rightarrow x$	Finite State Automata

A and B are single non-terminal nodes; α, β, and γ are arbitrary strings of terminal and non-terminal nodes; x is an arbitrary string of terminal symbols.

(Chomsky 1959); figure from (Jurafsky and Martin 2000:479)

Figure 6.2: The Chomsky Hierarchy

in much more detail in (Jurafsky and Martin 2000:478-490).

Human language, at least in the domain of syntax, can be shown not to be produced by a regular grammar. It is though to be mostly context free, though see Jurafsky and Martin (2000:489) for some examples of context-sensitive phenomena. Morphology, however, is considered to be more constrained and largely describable using a regular grammar, or a set of regular expressions.

One of the earliest finite-state morphological models was developed by Koskenniemi (1983) and further improved by Karttunen (1983). A somewhat different research background led to the development of two-sided finite-state networks (Kaplan and Kay 1981, Kaplan and Kay 1994).

A finite-state automaton (FSA) is a computational mechanism that moves between discrete *states*. The computational process start at the *initial state* and moves through a series of intermediate states to a *terminal state*, where it stops. The movement between

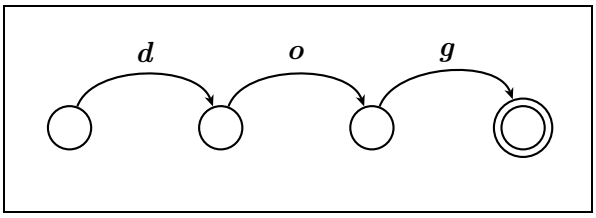


Figure 6.3: Finite-State Automaton for *dog*

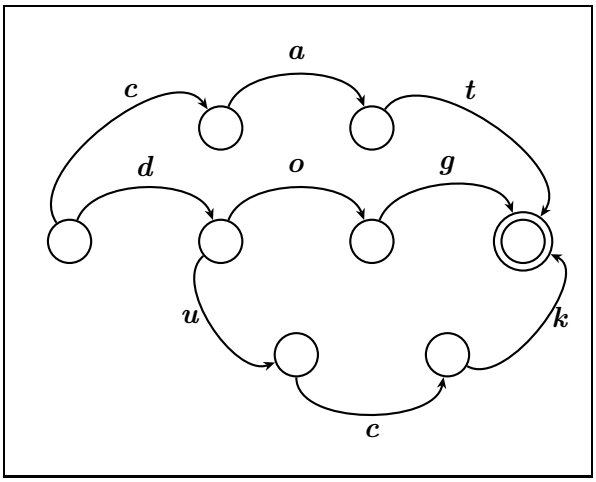


Figure 6.4: Finite-State Automaton for *dog*, *cat*, *duck*

states is done via *arcs*, which correspond to certain types of input. The finite-state diagrams, arcs are typically labeled with the corresponding input. To take an example of word recognition, movement between states can be precipitated by the input of specific letters (or phonemes), as in Figure 6.3. The finite-state automaton in this diagram recognizes the word *dog*, one letter at a time. If it receives the correct letters in the correct sequence, it arrives at the terminal state, indicated by the double circle. If it does not receive proper input (e.g *c*-*a*-*t*), it remains in one of the non-terminal state and thus *fails* to recognize the input.

A slightly more complex FSA in 6.4 recognizes three words: *dog*, *cat*, and *duck*. The first letters of *dog* and *duck* are shared, and the following input letter (*o* or *u*) makes the choice of direction.

In the domain of morphology, it is useful to represent not just words, but relations between morphosyntactic properties and word forms. For this purpose, finite-state transducers (FST) are useful. In addition to recognizing strings of symbols, and FST can

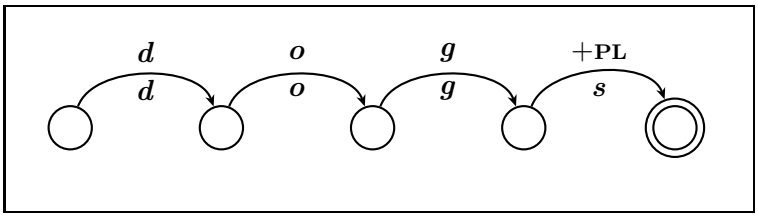


Figure 6.5: Finite-State Transducer for *dogs*

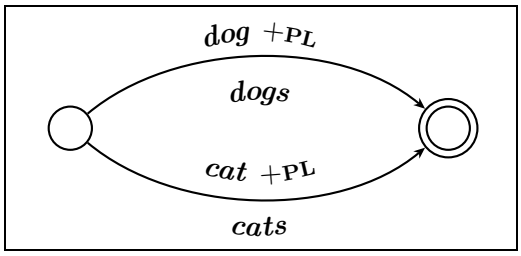


Figure 6.6: Finite-State Transducer for *dogs*, *cats*

produce an output string. In an FST, each arc contains a *mapping* between an input and an output string, so that the recognition of the *-s* in *dogs* produces *dog + PL*, as in Figure 6.5. The input symbols appear on the *upper side* of the arc, and the out appears on the *lower side*. Finite-state transducers are reversible and can be used to go from *dogs* to *dog + PL*, or from *dog + PL* to *dog*.

Modern FST applications provide sophisticated mechanisms for creating finite-state networks and transducers and are capable of handling very large networks. For simplicity of representation, individual letter-driven states can be abstracted away, so that each arc represents a full *path* from the initial state to the terminal state, as in Figure 6.6.

6.2.1 Standard Assumptions and Difficulties Presented by Georgian

Many contemporary approaches to computational morphology are based on, or can be easily translated into, finite-state networks (FSN). In such approaches, an arc in the FSN often corresponds to a phoneme or morpheme, and the recognition or generation of each arc advances the state in the network. Many approaches, including Beesley and Karttunen (2003), are implemented as two-way finite-state transducers (FST) in which each arc corresponds to a mapping of two elements, for example a phoneme and its phonetic realization or a morpheme and its meaning. As a result, FST morphology very often assumes

morpheme-level compositionality, the idea that the meaning of a word is compositionally made up from the meanings of its constituent morphemes. FST morphology has, for the most part, been applied to mainly agglutinative (i.e. concatenative) morphological systems like Finnish (under standard assumptions, at least), although there have been some recent applications to templatic morphology such as Arabic (Beesley 1996, Beesley and Karttunen, ch. 8).

As demonstrated in previous chapters, the assumption of morphemic compositionality does not contribute to the description of Georgian verbal morphology. The Georgian verb forms are made up of identifiable morphological elements (i.e. elements of form), but the meaning of these elements is not easily identifiable, and does not stay constant in different morphosyntactic contexts.

Briefly, these difficulties include the following.

- Meaning carried by a whole word form rather than by individual morphemes (Chapter 3)
- Lexical root alternations and suppletion (Section 3.3.6)
- Lexical class-dependent screeve formation, e.g. the endings in the Aorist (Section 3.3.6)
- Dependencies involving the formation of some screeves from that of others, e.g. the Imperfect from the Present (Section 3.3.6)
- The multiple exponence of agreement, i.e. the use of suffixes and prefixes simultaneously, and the simultaneous expression of subject and object agreement (Section 3.2.3)

As a reminder, the complexity of the distribution of morphological elements in Georgian is illustrated by preverbs, thematic suffixes, and tense endings here. The preverbs (a closed class of about 8) indicate perfective aspect and lexical derivations from roots, similar to verb prefixes in Slavic or German. The association of a verb with a particular preverb is lexical and must be memorized. A preverb appears on forms from the Future subgroup of series I, and on all forms of series II and III in transitive verbs. Table 6.1 demonstrates some of the lexically-dependent morphological elements, including several different preverbs (row 'Future').

Similarly, thematic suffixes form a closed class and are lexically associated with verb roots. They function as stem formants and distinguish inflectional classes. In transitive verbs, thematic suffixes appear in all series I forms. Their behavior in other series differs

	'Bring'	'Paint'	'Eat'
Present	i-**gh**-*eb*-s	**xat'**-*av*-s	**ch'am**-*ø*-s
Future	*c'amo*-i-**gh**-eb-s	*da*-**xat'**-av-s	*she*-**ch'am**-s
Aorist	c'amo-i-**gh**-*o*	da-**xat'**-*a*	she-**ch'am**-*a*
Perfect	c'amo-u-**gh**-*ia*	da-u-**xat'**-*av-s*	she-u-**ch'am**-*ia*
Masdar	c'amo-**gh**-eb-a	da-**xat'**-v-a	**ch'**-am-a

Table 6.1: Lexical Variation. Roots are in bold; lexically variable affixes are in italics.

by individual suffix: in series II, most suffixes disappear, though some seem to leave partial "traces" (rows 'Present' and 'Perfect' in Table 6.1).

The next source of semi-regular patterns comes from the inflectional endings in the individual tenses and the corresponding changes in some verb roots (row 'Aorist' in Table 6.1).

Finally, another verb form relevant for learners is the masdar, or verbal noun. The masdar may or may not include the preverb and/or some variation of the thematic suffix (last row in Table 6.1).

The linguistic analysis of Georgian verbal morphology suggested in the previous chapters relies on insights from Construction Grammar. Unfortunately, there are currently no computational implementations of CG capable of handling complex morphological systems. Bryant (2003) describes a constructional syntactic parser, based on general principles of chart parsing. However, this parser cannot yet handle morphological segmentation, and adapting it for Georgian would require a substantial revision.

Fortunately, FST tools for computational morphology have advanced to the point where they can handle some aspects of non-concatenative morphology. The next section briefly describes the approach in Beesley and Karttunen (2003) and what makes it a possible candidate for modeling at least a subset of Georgian verbal morphology.

6.2.2 Xerox Finite-State Morphology Tools

Beesley and Karttunen (2003) present the most recent set of tools for creating FST computational models. The book is accompanied by implementations of the two Xerox languages: *xfst* (designed for general finite-state manipulations) and *lexc* (designed more specifically for defining lexicons). Since my goal was to reproduce morphotactic rules

of word formation rather than the structure of the lexicon, *xfst* was used.

Xfst provides all of the basic commands for building up single or two-level finite-state networks (i.e. transducers), such as concatenation, intersection, etc. In addition, *xfst* has several built-in shortcuts that make network manipulation easier, such as various substitution commands. *Xfst* distinguishes between words of a natural language (composed of single characters) and multi-character symbols, used in our model to indicate morphosyntactic properties like person or number. Each completed arc in a finite-state network compiled using *xfst* represents a mapping between a set of morphosyntactic and semantic properties (on the upper side) and a full word form that realizes those properties (on the lower side), as in the examples above.

Another very useful feature of *xfst* is the ability to create scripts with several commands in a sequence. The later commands can operate on the output of earlier commands, and can thus create a cascade of finite-state transducers. *Xfst* also provides convenient ways of outputting all the words recognized by a given transducer, which proved very useful in the creation of the online reference (see Section 6.4). An updated version of *xfst* (Beesley and Karttunen 2006) also includes support for utf-8.

Finite-state technology is very good at generating and recognizing regular expressions and has been successfully applied to the modelling of agglutinative morphology like Finnish (cf. Beesley and Karttunen 2003) and Turkish (Oflazer and Inkelas 2006). However, it has a harder time capturing other features of natural language such as non-concatenative morphological structure. Augmented FST models have been adapted for Semitic languages (Beesley 1996 for Arabic, Cohen-Sygal and Wintner 2006 for Hebrew), but there have been no attempts to model non-compositional patterns in systems that appear to be agglutinative, like Georgian. The next section describes some adaptations that allow FST to handle many of the non-concatenative patterns of Georgian.

In addition, FST is not designed to represent a dynamic, living mental lexicon of an actual speaker. It does not provide any mechanisms for probabilistic decisions, or for recognition and generation of novel inflectional forms. More generally, FST provides a purely syntagmatic analysis and has no means of handling paradigmatic aspects of a structure, in which the interpretation of a given form depends on the form or interpretation of other forms. One such example is the formation of the Spanish indicative and subjunctive forms in Table 2.5 in Chapter 2.

6.3 Prototype Model of the Georgian Verb

6.3.1 General Idea

As argued above, Georgian verb morphology can be described as a series of patterns at various levels of regularity. Most of the patterns specify particular morphosyntactic or semantic properties of verb forms and the corresponding combinations of elements in the morphological templates. In the model proposed here, screeve formation is viewed as lexical or semi-regular, and pronominal agreement is viewed as completely regular. Screeve formation for different conjugation classes (transitive, unergative, unaccusative, and inverse) is fairly different in Georgian, and so each conjugation class is implemented as a separate network. The principles for composing each network are the same, however.

The model is implemented as a cascade of finite-state transducers; that is, several levels of FST networks such that the result of composing a lower-level network serves as input to a higher-level network. The levels correspond to the division of templatic patterns into completely lexical (Level 1) and semi-regular (Level 2). Level 3 contains completely regular patterns that apply to the results of both Level 1 and Level 2. The result of compiling Level 3 patterns is the full set of conjugations for the verbs whose lexical information is included in Level 1. The FST model can be used both for the generation of verbal inflections and for recognition of complete forms.

In general, the most specific or irregular information is contained at the lower levels. The higher levels, by contrast, contain defaults that apply if there is no more specific information. The verbs explicitly mentioned in the lexical level (Level 1) are representative examples of lexical classes, as posited by the linguistic analysis in Chapter 3. Through the use of diacritics and replacement algorithms, other verbs are matched to their lexical classes and included in the resulting network.

The main advantage of this implementation is in the separation of lexical, or irregular, verb formation patterns from the semi-regular or completely regular patterns. The initial input to the FST cascade includes only the necessary lexical information about each verb and verb class; the rest of the work is done by the computational model.

The model described here served as the basis for an online reference on Georgian verb conjugation, described in Section 6.4. This practical application underlies some of the specific choices in implementing the model. The prototype implementation of the model focuses on transitive verbs. However, there are obvious ways of extending the model to

apply to other verb classes.

The lexical classes of transitive verbs are given in Table 6.2, based on the classification by Melikishvili (2001). Classes are described here by exemplary forms that show the parameters of inflectional variation. Melikishvili's classification numbers classes in an absolute sequence (for a total of 63 paradigms), and also relative to 'diatheses', which largely correspond to valence/conjugation classes as described in Chapter 3. Thus, the classes in Table 6.2 are identified by both sets of numbers.

CLASS	PRESENT	FUTURE	AORIST2SG	AORIST3SG	OTHER	GLOSS
1(18)	*c'er*	*dac'er*	*dac'ere*	*dac'era*		'write'
2(19)	*k'ecav*	*dak'ecav*	*dak'ece*	*dak'eca*		'fold'
3(20)	*drek'*	*modrek'*	*modrik'e*	*modrik'a*		'bend'
4(21)	*ac'ven*	*daac'ven*	*daac'vine*	*daac'vina*		'weave'
5(22)	*arc'ev*	*daarc'ev*	*daarc'ie*	*daarc'ia*	Pres3Pl *c'rian*	'rock'
6(23)	*cri*	*gacri*	*gaceri*	*gacra*		'sift'
7(24)	*zrdi*	*gazrdi*	*gazarde*	*gazarda*		'raise'
8(25)	*agzavni*	*gaagzavni*	*gaagzavne*	*gaagzavna*		'send'
9(26)	*xat'av*	*daxat'av*	*daxat'e*	*daxat'a*	Perf *dauxat'avs*	'paint'
10(27)	*k'rav*	*shek'rav*	*shek'ari*	*shek'ra*	Perf *shek'ravs*	'tie'
11(28)	*asheneb*	*aasheneb*	*aashene*	*aashena*		'build'
12(29)	*akeb*	*sheakeb*	*sheake*	*sheako*		'praise'
13(30)	*ak'leb*	*daak'leb*	*daak'eli*	*daak'lo*		'decrease'
14(31)	*atbob*	*gaatbob*	*gaatbe*	*gaatbo*		'warm up'
15(32)	*nanob*	*(she)inaneb*	*(she)inane*	*(she)inana*		'regret'
15*(32)	*sesxulob*	*isesxeb*	*isesxe*	*isesxa*		'lend'
16(33)	*pq'rob*	*ipq'rob*	*sheipq'ari*	*sheipq'ro*		'arrest'
17(34)	*art'q'am*	*daart'qam*	*daart'q'i*	*daart'q'a*	Perf *daurt'q'ams*	'hit'

* - there is a typo in the original Melikishvili numbering.

Table 6.2: Classes of Transitive Verbs, from Melikishvili 2001

6.3.2 Level 1: the lexicon

The first level of the FST model contains lexically specific information. There are two separate networks. The first network contains information about the gloss and masdar or the verb stem.

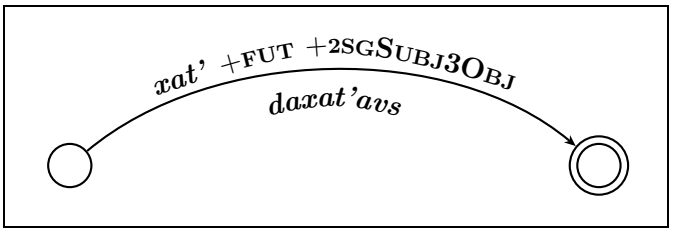


Figure 6.7: Finite-State Transducer for *daxat'avs* 'you will paint X

The second network contains several complete word forms for each verb stem, providing all the lexically-specific information needed to infer the rest of the inflections. For the most regular verbs, these are:

- Present screeve, no overt agreement (corresponds to 2Sg Subject, 3Sg Object)
- Future screeve, no overt agreement
- Aorist screeve, no overt agreement
- Aorist, 3Sg Subject, no overt object agreement
- Aorist Subjunctive

Some verbs need additional forms in order to describe their paradigms.

- Present screeve, 3Pl Subject (most verbs have the ending *-en*, but some end in *-ian*)
- Perfect screeve

The inflected forms are represented as two-level finite-state arcs, with the verb stem and morphosyntactic properties on the upper side, and the inflected word on the lower side, as in Figure 6.7. The purpose of the stem is to uniquely identify each verb. Verb roots in Georgian are often very short and ambiguous; therefore a combination of the verb root plus thematic suffix was used. In some cases, even this combination would be insufficient to identify the verb uniquely; in such cases, the preverb may be necessary as well. It is only important that the verb stem can be uniquely matched in the network containing glosses; thus the stem has no theoretical significance in this model.

Another challenge is posed by the non-concatenative nature of verb agreement. Recall from Chapter 3 that verb agreement is realized by a pre-stem affix and a final suffix. Since many of the word forms in Level 1 contain preverbs, the agreement affix would need to be infixed into the verb form at a later level. Beesley and Karttunen provide some fairly complex mechanisms for doing phonological infixation in FST. However, the fixed position of the agreement affixes in the Georgian verb template allows for a much simpler solution.

The forms on Level 1 contain a place holder "+Agr1" for the prefixal agreement marker (Figure 6.8), which is replaced by the appropriate marker in the later levels.

The Level 1 network is produced via scripts from a table of verb forms containing only the necessary lexical information. Redundancy in human input is thus minimized.

6.3.3 Level 2: semi-regular patterns

The purpose of Level 2 is to compile inflectional forms that are dependent on other forms (introduced in Level 1), and to provide default inflections for regular screeve formation patterns.

An example of the first case is the Conditional screeve, formed predictably from the Future screeve. This represents a parasitic or Priscianic formation, in the sense of Matthews (1972, 1991). The FST algorithm is as follows:

- Compile a network consisting of Future forms.
- Add the appropriate inflectional suffixes (-di for 1st and 2nd person subject, -da for 3rd person subject).
- Replace the screeve property "+Fut" with "+Cond".
- Add the inflectional properties where needed.

The replacement of screeve properties is done using the "substitute symbol" command in *xfst*; other operations are performed using simple concatenation commands.

An example of the second case is the Present 3PlSubj suffix, which is *-en* for most transitive verbs, but *-ian* for a few others (see Fig. 6.10). *Xfst* provides a simplified feature unification mechanism called *flag diacritics*. Using these flags, we specify exceptional forms in Level 1, so that default inflections do not apply to them in Level 2.

All of the patterns defined at Level 2 are then compiled into a single network, which serves as input to Level 3.

6.3.4 Level 3: Regular Patterns

The purpose of Level 3 is to affix regular inflection, namely subject and object agreement. As described in Chapter 3, agreement in Georgian is expressed via a combination of a pre-stem affix and a suffix, which are best thought of as attaching simultaneously and working in tandem to express both subject and object agreement. Thus the compilation of Level 3 consists of several steps, each of which corresponds to a paradigm cell.

```
# Level 1 – Lexicon
# +Agr1 is the place holder for agreement

[[xat' +Pres .x. +Agr1 xat'] |                    Present-tense form, no agreement
[xat' +Fut .x. da +Agr1 xat'] |                   Future-tense form, no agreement
[xat' +Aor .x. da +Agr1 xat'e] |                  Aorist form, no agreement
[xat' +Aor +3SgSubj .x. da +Agr1 xat'a] |         Aorist 3sgSubj form
[xat' +Aor +3PlSubj .x. da +Agr1 xat'es] |        Aorist 3plSubj form
[xat' +Perf .x. da +Agr1 xat'avs]]                Perfect form
# Level 2 – Derivative forms
# Conditional                                       Extract future forms,
regex [Level1.i .o. [Level1.u & [?* +Fut ?*]]].i    Extract future forms
      [+Cond: di ];                                 Add Conditional endings to Future
substitute symbol 0 for +Fut                        Change properties accordingly
define Cond12;
# Level 3 – Agreement
# 1SgSubj 2SgObj
# Extract only forms not already marked for 3rd person subjects
regex [Level2.i .o. [Level2.u - [Level2.u & [?* +3SgSubj ?*]] -
[Level2.u & [?* +3PlSubj ?*]]]].i;
substitute symbol m for +Agr1
define temp
regex temp [[+1SgSubj +2SgObj]:0];
define 1SgSubj2SgObj;
```

Figure 6.8: Simplified FST script

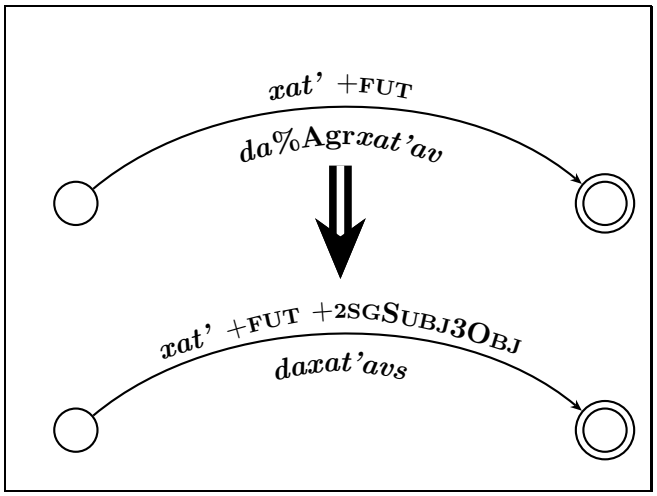


Figure 6.9: Adding agreement affixes

In each step, all of the word forms from Level 2 are taken as input. The place holder for the pre-stem agreement affix is then replaced by the appropriate prefix (in some cases, this is null), and the appropriate suffix is attached at the end, as in Figure 6.9. The resulting networks are then compiled into a single network.

The only difficulty at this level arises when dealing with the "inverted" screeves (Perfect and Pluperfect). As demonstrated in Chapter 3, the morphological agreement in these screeves is sensitive to the case-marking of the nominal arguments, which is the reverse of the regular pattern. However, the composition of the agreement paradigm is sensitive to the semantic roles played by the arguments: plural number agreement is still triggered by the semantic agent. In this case, the computational implementation was motivated by the practical application of the model to the online reference. A separate set of paradigm cells was created for the inverted tenses, interpreting the properties "Subject" and "Object" as semantic. The resulting FST network thus shows no relation between inverted and non-inverted forms (i.e. it does not capture the generalization behind inversion). Such an interpretation was sufficient for the purposes of the conjugation reference. However, the model could easily be amended to incorporate a different analysis of inversion that relies on the distinction between semantic and syntactic arguments.

Lev. 1	*paint*+Pres	*paint*+Aor	*open*+PresPl
	*xat'-**av***	***da**-xat'-**a***	*xsn-**ian***
Lev. 2	*paint*+Past+3Sg	*paint*+Pres+3Pl	default
	*xat'-av-**da***	*xat'-av-**en***	overridden
Lev. 3	*paint*+3PlSubj+1SgObj		*open*+3PlSubj+1SgObj
	***m**-xat'-av-en*		***m**-xsn-ian*

Figure 6.10: Verbs 'paint' and 'open' at three levels of the model. New information contributed by each form is in bold.

6.3.5 Treatment of Lexical Classes

Georgian verbs are divided into lexical classes based on inflectional patterns in different screeves and series. Lexical classes are subdivisions of the four conjugation classes (transitive, unaccusative, unergative, and indirect) identified by Harris (1981) and discussed in more detail in chapter 3. A subset of the transitive verb classes with their exemplar representatives is in Table 6.2 above.

The differences in formations include different thematic suffixes, preverbs, Aorist screeve endings, vowel changes in the stems and some other exceptional behavior indicated in the "Other" column. Melikishvili's classification aims for "maximal generality" of patterns: once the class of a verb is identified, its formation is completely predictable using its stem. The only lexical component of each verb is the preverb.

However, the boundary between what is lexical and what is determined by inflectional class is not precise. For example, it is possible that the stem vowel changes are memorized on a verb-by-verb basis and are not generalized into inflectional class patterns, because such changes are not frequent. With that assumption in mind, the present and Aorist stem for each individual verb have to be memorized; and we can merge the verb classes 1, 3, and 4; 7 and 8; 9 and 10 above, reducing the total number of classes from 18 to 14. However, the number of classes is in some sense immaterial, as long as it is significantly smaller than the overall number of verbs.

The input to Level 1 contains a representative for each lexical class, supplied with a diacritic feature indicating the class number. Other verbs that belong to those classes could, in principle, be inputted along with the class number, and the FST model could substitute the appropriate roots in the process of compiling the networks. However, there are several challenges to this straightforward implementation:

First, verbs belonging to the same class may have different preverbs as well as different roots, thus complicating the substitution. Second, for many verbs, screeve formation involves stem alternations such as syncope or vowel epenthesis, again complicating straightforward substitution. Third, suppletion is also quite common in Georgian, requiring completely different stems for different screeves.

As a result, even for a verb whose lexical class is known, several pieces of information must be supplied to infer the complete inflectional paradigm. The FST substitution mechanisms are fairly restricted, and so the compilation of new verbs is currently done using Java scripts performing simple string manipulations. Such an implementation still makes use of the division into lexical classes. The scripts make non-example verbs look like example verbs in Level 1 of the FST network by creating the necessary inflected forms, but the human input to the scripts need only include the information necessary to identify the lexical class of the verb. Future improvements to the computational model may include a more efficient method of identifying lexical classes within FST itself.

Another issue is the psychological reality of the lexical classes. The pilot survey of morphological productivity conducted with adult speakers of Georgian (described in Section 3.3.7) suggests that speakers conjugating nonce verbs sometimes rely more on frequent inflectional patterns than on a rule-based comparison with existing verbs based on morphophonological similarities with the nonce verbs. Such a reliance on frequency is not reflected in Melikishvili's classification. The computational model proposed here takes a small step in this direction by relying on frequent verbs as example paradigms. However, the model does not have any built-in way to accommodate the relative frequency of different inflectional patterns. The concluding section suggests some possible improvements for the future.

6.3.6 Case Frames

As described in Chapter 3, another complicating feature of the Georgian verb is the variability of case-marking patterns for the verb's arguments. For the purposes of the online conjugation reference, it was necessary to present the case-marking information with each verb. Fortunately, the case marking patterns depend almost entirely on the conjugation class and TAM series of the verb. Since the goal of the online reference is to describe the morphosyntactic patterns of Georgian, it was sufficient to simply mention the case-marking pattern for each verb type. If the purpose of the morphological transducer is to supplement

a syntactic parser, the case-marking information could be represented as a feature structure and associated with each verb type. Meurer (1999) describes one such proposal using LFG-style feature structures, so that the output of the morphological processor can be used by a larger LFG parser. Given the syntactic analysis of Georgian morphosyntax proposed in chapter 3 of this dissertation, a constructional feature-structure representation could be equally plausible. However, since a syntactic parser was not the goal of the computational model proposed here, the precise orthographic representation of case frames is not crucial.

6.3.7 Previous Approaches

A few computational models of Georgian verbs have been proposed in the literature, although none are described in great detail or currently available to the public. Datukishvili (1997) describes a model that combines various morphemes using templatic formulas, thus accounting for the verb template described in Chapter 3. However, based on the short available description, this model would produce *possible* combinations of morphs for Georgian and has not way of describing lexical variation and predicting co-occurrence constraints such as the ones exemplified in Table 6.1. The model proposed by Margvelani et al. (1997) appears to be similar, largely compatible with a finite-state representation, but again it is not clear how lexical variation is constrained. Tandashvili (1999) focuses mainly on the expression of different syntactic frames rather than morphotactic patterns. Finally, Meurer (1999) relies on a scanned-in dictionary of Georgian verb forms (Tschenkéli 1960-1974) and, due to copyright restrictions, this model has not yet been described in detail or made publicly available. None of these approaches, as far as one can tell from the descriptions, expresses form interdependencies such as those between Future and Conditional, or Aorist and Perfect.

6.3.8 Summary

The computational model presented here accommodates many properties of Georgian verbal conjugation that make it challenging: the templatic structure of the verb forms; the non-concatenative nature of word meaning construction; the large number of irregular and semi-regular word formation patterns; and the interaction between word formation and case marking on the verb's arguments. The model crucially relies on classification of verbs into lexical classes with example paradigms for each class. The two-level mappings inherent

in FST mean that the model can be used for generation as well as recognition.

6.4 Practical Application: An Online Reference

6.4.1 Purpose

The computational model described here serves as the basis for an online reference on Georgian verb conjugation. The goal of the online reference is to aid the learners of Georgian in several ways:

- Complete conjugation tables for 200 frequently-used verbs.
- The verb database can be searched using any verb form or its English translation.
- For many verb forms, real-life examples from the Internet, audio and video sources are provided, along with translations.
- Several types of exercises are available on the website; answers are automatically checked for correctness.

The online reference is meant as an addition to the classroom or self-study using a textbook, such as (Kurtsikidze forthcoming).

6.4.2 Website design

The website is divided into four sections: "Verb Conjugation," "Examples," "Exercises," and "Resources." The section on verb conjugation is the core of the reference tool. It provides complete tables of verb conjugations, accessible through browsing by individual verb (in Georgian or in English), or by searching. The conjugated forms are produced using the FST model described in the previous section; the forms are then automatically inputted into a MySQL database and displayed on the website using PHP. In addition to displaying verb forms, the site allows the user to search for a given verb form, using the recognition capabilities of the FST network. This search capacity demonstrates a major advantage of online resources over print.

Many of the verb forms are accompanied by hand-picked examples of usage from print sources (mainly online newspapers and chat rooms), audio (from recorded naturalistic dialogues), and movie clips. The examples are provided as complete sentences and short paragraphs; translations are available for all examples. Audio and video examples are

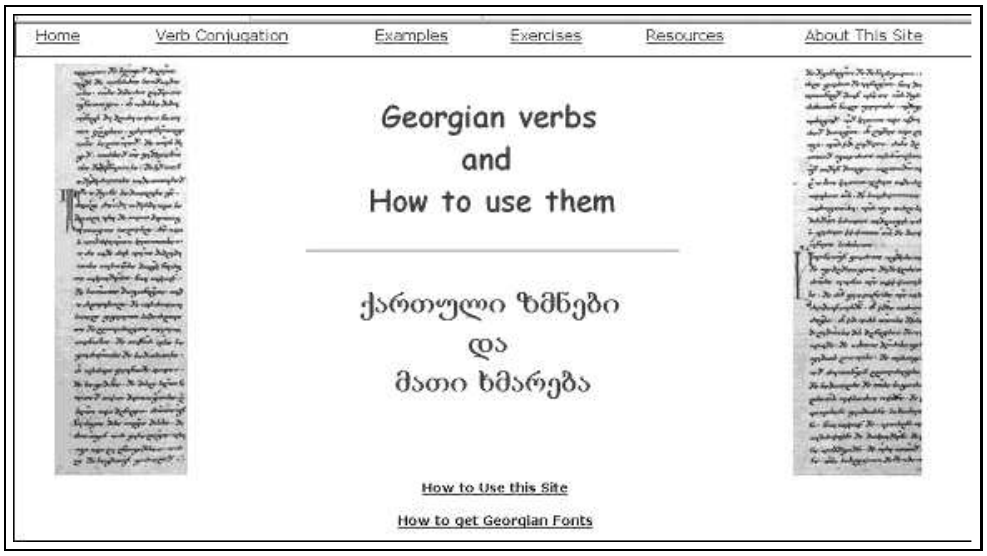


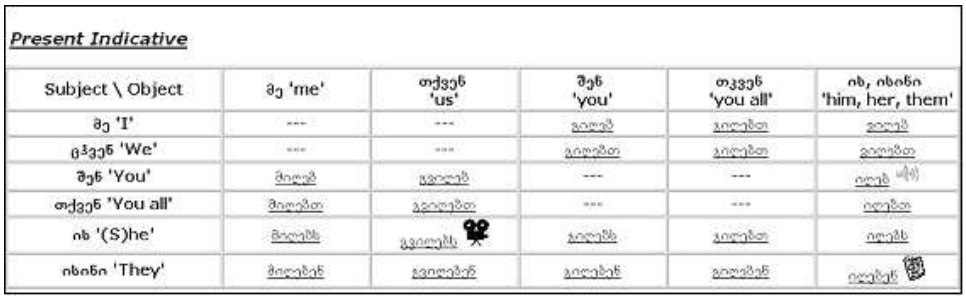
Figure 6.11: Online reference: First page and division into sections

likewise accompanied by transcriptions and translations.[2]

The "Examples" section of the website provides a different way to access the print, audio, and video examples. This can be done through browsing by verb or by searching (again, in Georgian or in English). The "Exercises" section contains several different types of exercises to provide additional practice for using and conjugating verbs. Many of the exercises are generated based on the conjugated forms or the hand-picked examples, and so the correctness of the answers can be checked automatically. Finally, the "Resources" section contains links to various online and bibliographical resources about Georgian, as well as technical suggestions for using Georgian fonts.

Figure 6.11 is a screen shot of the front page of the reference, showing access to the various divisions. Figure 6.12 is an example of a screeve table with links to print, audio, and video examples. The website will be operational in the summer of 2006; anyone interested in using it should contact this author or search for "online reference for Georgian verb conjugations" on the Web.

[2]I am very grateful to Vakhtang Chikovani for finding and translating the examples.



Present Indicative

Subject \ Object	მე 'me'	ჩვენ 'us'	შენ 'you'	თქვენ 'you all'	ის, ისინი 'him, her, them'
მე 'I'	---	---	გიღებ	გიღებთ	გიღებ
ჩვენ 'We'	---	---	გიღებთ	გიღებთ	გიღებთ
შენ 'You'	მიღებ	გგიღებ	---	---	იღებ
თქვენ 'You all'	მიღებთ	გგიღებთ	---	---	იღებთ
ის '(S)he'	მიღებს	გვიღებს	გიღებს	გიღებთ	იღებს
ისინი 'They'	მიღებენ	გვიღებენ	გიღებენ	გიღებენ	იღებენ

Figure 6.12: Online reference: Screeve conjugation table with links to examples

6.5 Conclusions and Further Work

This chapter represents a prototype model Georgian verbal morphology using easily available, off-the-shelf technology such as FST. Using some adaptations to accommodate the templatic and non-compositional structure of the Georgian verbs, I was able to make significant progress and express some non-compositional inflectional patterns such as parasitic formation. The model provides a convenient method for representing a subset of the existing lexicon for computational applications such as parsing or generation, and serves as a useful basis for an online learner's reference. At the same time, this exercise has demonstrated some fundamental limitations of finite-state technology for modeling the flow of information in inflectional systems, such as the inability to express paradigmatic interdependencies.

FST provides no way to incorporate frequency information about the Georgian lexicon, and in general is not an accurate model for how verbs are learned. Unfortunately, creating a statistically sensitive model of the Georgian lexicon is not currently an easy proposition as there are no available corpora of Georgian, and no immediate ways of obtaining statistical distributions. However, it is hoped that this model and the online reference and collection of examples can serve as the basis for the creation of a corpus of spoken Georgian. Information collected in the corpus can then be used to inform and improve future computational models. In the near future, the existing model will be enriched with more verb types and more inflectional parameters (such as the use of pre-radical vowels and productive passivization and causativization processes).

Chapter 7

Extensions and Implications

This chapter sums up the results of the previous chapters and examines possible extensions and implications of the Constructional Morphology model proposed in this dissertation.

7.1 Constructional Morphology: Summary of Arguments

The previous chapters examined several phenomena dealing with Georgian verbs and argued for a constructional representation of the syntax, morphology, and the syntax-morphology interface. The main arguments for this approach rely on the following facts. First, the formation of Georgian verbs falls into several patterns at various degrees of regularity. Among the most irregular is the choice of screeve endings in the Aorist, stem suppletion, and stem ablaut. A more regular pattern concerns the behavior of preverbs and thematic suffixes in various screeves. And the most regular part of Georgian verb inflection is the subject-object agreement, which does not depend on individual verbs. These patterns are most stable if they are described relative to a conjunction of conjugation class and TAM series.

Second, the inflectional choices in some forms of a verb tend to predict choices in other forms through a set of implicational relationships. For example, the choice of Aorist ending tends to predict the formation of the Pluperfect. This predictability is purely morphological, as the predicted form is often not derived, either phonologically or semantically, from the predicting form. The sets of predictability relations can be described as lexical classes, such that all Georgian verbs fall into one of the classes, and a few leading forms

identify a verb as belonging to a given class.

Second, although the elements of form that make up Georgian verbs are easily extractable from the full word forms, their function crucially depends on other properties of the verb form in which they occur. Pronominal agreement markers present one such example: although the individual elements are clearly identifiable, their function is best described in tandem with other agreement elements on the verb (i.e. prefixal markers together with suffixal markers), and in contrast to other forms in a given verb's paradigm. In addition, the connection between agreement markers and thematic roles depends crucially on the conjugation class and series of the verb, via a set of mapping rules. The function of pre-radical vowels represents a more complex example of the same phenomenon. The pre-radical vowels, easily identifiable as occupying a particular slot in the verb template, nevertheless have an astoundingly wide array of functions, from the core subsystem of *version*, or participant affectedness marking, to a set of voice and valence-related alternations. However, all of these functions can be described predictably on the basis of the series and conjugation class of a verb.

In summary, a description of Georgian verbal morphosyntax that aims to capture all of the relevant generalizations inherent in the data must rely on a set of simultaneous morphological, syntactic, and semantic constraints that predict verb formation from the series/screeve and conjugation class. The series is the most stable a *constructional* unit, around which all regularities can be expressed. In a "top-down" fashion, the series constructions for particular conjugation classes dictate the properties of the full words that instantiate them. In such a description, the words themselves consist of an inventory of purely morphological/phonological elements, or morphs, which serve as formal building blocks for words, but whose function is determined from the top by the whole word. By virtue of consistent association with certain constructions, such morphs may have held unitary compositional meanings at times in history (as suggested, for example, the extension of middle-voice marking vowel *i-* from unergative/medioactive verbs to reflexive subjective version), but their synchronic description cannot assign such individual meanings separate from the constructions in which they participate.

The intuition that larger units may determine meaning and select for their component parts is an old one. In both syntax and morphology, it dates to the pre-Bloomfieldian era and the neogrammarians. Recent work in psycholinguistics suggests that the top-down methods are, indeed, employed both by adults and by children learning language, and that

syntactic and morphological patterns are extended analogically from individual instances. Modern Construction Grammars have revived this intuition for syntax and semantics, and this dissertation is a step towards formalizing it for morphology and morphosyntax.

In the rest of this chapter, Section 7.2 examines several related classes of phenomena, in languages unrelated to Georgian, that demonstrate the relevance of construction-based morphology on a wider scale, and Section 7.3 provides general suggestions for detecting constructional patterns in morphology and draws some general conclusions.

7.2 Constructional Morphology Outside of Georgian

This section argues for the relevance of the construction-based view of morphology to some further phenomena outside Georgian. Although explicitly construction-based phenomena can be found in almost any aspect of syntax or morphology (as witnessed by the growing body of literature in this domain), for the sake of cohesiveness I will focus on phenomena most closely resembling those patterns described for Georgian.

One class of such phenomena involves formal elements that, for various historical reasons, have diverse synchronic functions in different syntactic contexts and in combination with different lexical elements.[1] Reflexive clitics/suffixes in Slavic illustrate this problem quite clearly and are discussed in Section 7.2.1; the major point is that synchronically, they can only be described as morphological material recruited by larger constructions or lexicalized in combination with some verbs, but do not have unitary, invariant meaning that compositionally contributes to the meanings of these larger constructions.

The diverse function of reflexive suffixes is part of a larger class of phenomena with similar properties, namely voice alternations. The so-called "middle" voice, best known from Indo-European studies, generally tends to have a wide variety of uses, as described in Section 7.2.2. Here, a certain class of syntactic and semantic contexts, best thought of as constructions, requires the use of middle-voice verb forms. The range of such functions overlaps to a large extent with the range of functions of Georgian version vowels.

Other voice alternations exhibit properties that are problematic for completely lexical and derivational accounts. For instance, Tagalog and other Philippine-type languages demonstrate clause-level patterning for voice that cannot be attributed to lexical properties

[1] Zwicky (2001) suggests that elements of form, such as stems, are usually independent of elements of meaning, and their multifunctionality is the general case rather than an exception.

of specific verbs (Section 7.2.3). Section 7.2.4 also briefly discusses active and passive voice in Coast Salish languages.[2]

7.2.1 Slavic Reflexives

One example of a formal element that has a variety of functions is the Slavic reflexive suffix/clitic *se/sja*. In addition to the standard syntactic function of indicating argument co-reference, reflexive suffixes are found in verbs that cannot formally be said to have reflexive syntax. Some of the functions of the reflexive suffix in Russian are illustrated in (178).

(178) Synchronic functions in Russian (from Timberlake 2004):

1. "True" reflexives: *myt'sja* 'wash oneself', *brit'sja* 'shave oneself' as contrasted with *myt'* 'wash something', *brit'* 'shave something'
2. Reflexivum tantum (no variant without the suffix): *bojat'sja* 'be afraid (of someone)', *slushat'sja* 'obey (someone)'
3. Reciprocal reflexives: *drat'sja* 'fight', *mirit'sja* 'make up'; not possible without the suffix
4. Habitual reflexives: *sobaka kusaetsja* 'the dog bites'
5. Phenomenological reflexives: *beleetsja* 'it whites (something white is seen)'
6. Model impersonal reflexives: *mne ne spitsja* 'I don't feel like sleeping', *mne ne rabotaetsja* 'I don't feel like working'
7. Quantifying reflexives: *naglotalsja* 'he swallowed a lot/enough', *doigralsja* 'he has played until (something bad happened)'
8. Intransitivized reflexives: *nachatsja* 'to start', *izmenitsja* 'to change'
9. Reflexive passives: *konfiskovyvalos'* 'it was confiscated'

There have been many suggestions for how these meanings may relate historically or synchronically at some deep level (Gerritsen 1990, Israeli 1997, Siewierska 1988, Švedova 1980, Vinogradov 1947), but many of the different functions seem unrelated synchronically.

[2]Another category that is often described in conjunction with / opposition to middles, reflexives, and passives, are impersonal constructions (cf. Blevins 2003).

The same kind of functional diversity is observed for Romance reflexives and, more generally, reflexives cross-linguistically (Geniušiene (1987) identifies over fifteen distinct functions).

From a constructional perspective, the different construction types identified in (178) can be said to select for the 'reflexive' suffix, without necessarily assuming that the suffix contributes invariant meaning to the whole word. Thus, historical connections may have given rise to the unified form, but the meaning need not be connected to the form in a one-to-one fashion. The next section discusses a category that is often correlated with the reflexives, namely middle voice.

7.2.2 Middle Voice

The parallels between Georgian pre-radical vowels and middle voice has already been discussed in Section 5.7.1. This section examines the functions of middle voice in Indo-European and some other languages mentioned in Klaiman (1991) and Kemmer (1994). I will suggest that middle voice represents a type of construction with a central, or prototypical, semantics, which integrates into the morphosyntax of particular languages and can be extended to cover related functions.

Klaiman (1991) distinguishes between two types of voice systems attested cross-linguistically. The so-called 'derived' voice has an alternation between active and passive forms, as in English. The so-called 'basic' voice systems, claimed to be more common in Latin and Greek, alternate between active and middle voice. According to Lyons (1968:373), "The implications of the middle (when it is in opposition with the active) are that the 'action' or 'state' affects the subject of the verb or his interests." In active voice, by contrast, the action primarily affects the patient/direct object. While the passive is often said to be derived from, or secondary to the active, the active and middle in an active/middle alternation are equal partners. Many verbs in Indo-European can inflect for either voice, as the Sanskrit example in (179) demonstrates. A parallel distinction, traditionally termed STRONG vs. WEAK, is found in the Dravidian language Tamil (180).

(179) Sanskrit middle (Klaiman 1991:24)

a. *Devadattaḥ* *kaṭaṃ* *karoti*
Devadatta-NOM mat-ACC makes-SG **ACTIVE**
'Devadatta makes a mat.'

b. *Devadattaḥ kaṭaṃ kurute*
Devadatta-NOM mat-ACC makes-SG **MIDDLE**
'Devadatta makes (himself) a mat.'

(180) Tamil middle (Klaiman 1991:25)

a. *Kuẓantai enn̲ai utai -kki̲r -atu*
child-NOM me-ACC kick present **STRONG** SG NEUT
'The child is kicking me.'

b. *Kuẓantai kālai utai -ki̲r -atu*
child-NOM leg-ACC kick present **WEAK** SG NEUT
'The child is kicking its legs (in the air).'

In the Niger-Congo language Fula, many verbs have paradigmatic oppositions between all three voice types: active, middle, and passive (181).

(181) Fule middles (Klaiman 1991:26)

a. *'o ɓorn -ii mo ŋgapalewol*
he dress past **ACTIVE** him gown
'He dressed him in a gown.'

b. *'o ɓorn -ake mo ŋgapalewol*
he dress past **MIDDLE** him gown
'He dressed him in a gown.'

c. *'o ɓorn -aama mo ŋgapalewol*
he dress past **PASSIVE** him gown
'He dressed him in a gown.'

Kemmer (1994) provides a crosslinguistically-based summary of situation types in which middle voice is typically used, given in (182).

(182) Middle situation types (Kemmer 1994:182-3)
(1) Grooming or body care
(2) Nontranslational motion
(3) Change in body posture
(4) Translational motion
(5) Naturally reciprocal events
(6) Indirect middle
(7) Emotion middle
(8) Emotive speech actions
(9) Cognition middle
(10) Spontaneous events

Klaiman (1991) surveys the voice systems of Fula, Tamil, and Indo-European languages with an aim to determine the precise function and distribution of middle voice verbs. Consistent with Kemmer's classification, the survey indicates a significant overlap in the syntactic and semantic functions of middles in different (unrelated) languages. First, the middle tends to correlate with "various kinds of non-eventuality, e.g. with atelic, nonpunctual, and/or irrealis temporomodal categories of the verb." In addition, when contrasted with the active, the middle tends to denote reduced valence or detransitivization, as summarized below.

(a). In every surveyed system, middle-inflecting verbs demonstrate a greater statistical tendency than active-inflective verbs to be intransitive.

(b). Neuter meaning is expressed by middle verbs in every surveyed system. Generally, such verbs are intransitive. In the majority of systems (Fula and IE), neuters comprise a subclass of middle-only inflecting verbs; while elsewhere (i.e. in Tamil), they comprise middle alternates of alternating active/middle verbs.

(c). Each surveyed system evinces a distinction among three sub-classes of lexical verbs: active-only (non-middle inflecting), middle-only (non-active inflecting), and alernating active/ middle. Among these, the basic valence of verbs of the third class is nearly always transitive, making intransitivity more characteristic of verbs which eihter are incapable of forming middles, or verbs whose primary voice is middle rather than active.

(Klaiman 1991:105)

Commenting on semantic properties of the middle, Klaiman notes, "[...] either as a characteristic voice of middle-only as contrasted with active-only verbs, or in opposition to the ative voice for active/middle inflecting verbs, the middle implicates the logical subject's affectedness." The association between middles and intransitivity is not complete and is motivated, but not determined, by the semantics. In addition, one voice cannot be said to derive from another, as there are many verbs that inflect for just the active or just the middle voice:

(a). Every basic voice system is organized into classes of active-only, middle-only, and alternating active/middle inflective lexical verbs [...].

(b). Among the three verbal classes listed in (a), in every basic voice system surveyed, the alternating active/middle class is the largest.

> (c). The two nonalternating verbal classes of a basic voice system are characteristically of unequal size. One of the two classes [...] is larger than the other. [...]
> item[(d).] While the middle demonstrates some affinity with intransitivity [...], the active/middle opposition criss-crosses distinctions of lexical transitivity and intransitivity.
>
> (Klaiman 1991:106)

Some examples of non-alternating verbs from Fula are in (183) and (184).

(183) Active-only verbs in Fula (Klaiman 1991:56)

a. *'o holl -ii mo ɗum*
he show General Past **ACTIVE** him it
'He showed it to him.'

b. *'o nodd -ii mo*
he call General Past **ACTIVE** him
'He called him.'

c. *'o maay -ii*
he die General Past **ACTIVE**
'He died.'

(184) Middle-only verbs in Fula (Klaiman 1991:56)

a. *'o ug -ake mo nde*
he throw General Past **MIDDLE** him it
'He showed it to him.'

b. *'o sal -ake nde*
he refuse General Past **MIDDLE** it
'He refused it.'

c. *'o dar -ake*
he stop General Past **MIDDLE**
'He stopped.'

Thus, both Klaiman (1991) and Kemmer (1994) argue that middle voice is essentially a semantic category, whose syntactic properties are secondary to its semantic properties. However, it does tend to occur in fairly specific syntactic and morphological constructions in the particular languages that exhibit it. Thus, for any given language, the middle voice is best described as a combination of semantic and formal constraints. It

appears that the 'meaning' of middle cannot be described as tied to any specific morphological element, but rather as a set of morphosyntactic constructions in paradigmatic contrasts with other constructions denoting non-middle voice. This property can find natural expression in a constructional account that can select morphological properties top-down based on the larger construction type.

The suggestions of a common, or prototypical, semantics of the middle also find easy expression in a constructional account. Similar to the German applicative pattern described by Michaelis and Ruppenhofer (2001), middle constructions seem to form a family that extends from the prototypical expression of subject-affectedness (perhaps prototypically intransitive) to other situation types that are associated with it. At the same time, the specific subset of situation types that uses middle voice in individual languages must be specified separately, as they are motivated, but not determined, by the more abstract general meaning.

The alternation between active and middle constructions, and the fact that verbs may participate in any one or both of them, is a particularly nice example of how a constructional account may work. Just as with Georgian version, the active and middle construction impose sets of semantic and syntactic constraints, and verbs with compatible semantics may unify with either or both types of constructions. The generalizations about middle- or active-voice semantics are expressed by the constructions, but there is no need to proliferate lexical entries for verbs to indicate which voice type they can participate in.

A constructional account becomes even more useful when comparing middle voice cross-linguistically. Since there is no single set of morphological or syntactic properties that mark middle constructions, the *only* way to identify and contrast them in different languages is by looking at whole constructional families, identifying semantic and syntactic commonalities while allowing for the unique properties of each language. Purely formal, or syntactic, criteria, would not allow for the full range of middle constructions to be identified cross-linguistically, as the overlap between 'middle' and intransitivity is only partial.

The middle has sometimes been identified as a subtype or supertype of the reflexive category (Woodcock 1959, Croft et al. 1987). Both Klaiman (1991) and Kemmer (1994) examine this relationship and recognize a significant overlap in the functions of middles and reflexives. However, they suggest that the two categories are nevertheless distinct. Klaiman proposes that the middle and reflexive have separate prototypes: while middles mark subject affectedness, the "prototypical meaning of reflexives is [...] the structural

reflexive, involving coincidence in the referents of verbally assigned core nominal positions." (Klaiman 1991:105). In other words, while the middle is at its core a semantic category, the reflexive is more easily identified by its syntactic functions. This type of categorization is, again, constructional in nature: if the two categories are identified at different levels of grammar, they can only be compared via the complex of semantic and structural features that embody them.

Another advantage of a constructional approach is the easy separation it allows between describing patterns of transitivity (syntactically-based) and patterns of middle vs. active voice (semantically-based). The two sets of distinctions appear to be somewhat orthogonal. In a lexicalist account, one would have to specify both the valency of a given verb and the types of voice alternations in which the verb may participate. This could potentially lead to a multiplication of verb entries. On a constructional account, however, transitivity and voice are classified in parallel, as generalizations over existing exemplar constructions. The generalizations concerning transitivity constitute one network of patterns (see Goldberg 1995 for arguments in favor of argument structures as constructions), and generalizations concerning middle vs. active voice are part of a separate network.

The next section examines some more cases of voice/valence alternations that exhibit constructional properties, namely the determination of form and meaning by a unit larger than a morpheme or single word.

7.2.3 Voice Alternations in Philippine Languages

Voice or 'focus' systems in Philippine languages such as Tagalog, Cebuano, or Kapampangan have been a subject of controversy in the linguistic literature for a very long time. One major issue concerns the classification of these languages as having one of the major orientations of grammatical relations: active-passive (Bloomfield 1917, Blake 1925, Bell 1983, Givón 1979, Kroeger 1993) or ergative-absolutive (Cena 1977, De Guzman 1979, Gerdts 1988, Mithun 1994), or being of a separate 'topic-focus' orientation that does not depend on traditional relations such as subject and object (Ramos 1974, Shibatani 1988). Another, related, major question concerns the status of grammatical relations in these languages, namely how to determine the subject in a given sentence.

The basic pattern of alternations is demonstrated for Tagalog in (185). All nominals in sentences are marked by one of three classes of preposed particles; the classes of

particles correspond to case paradigms. The first class, represented by *ng* in the example below, marks what Foley (forthcoming) calls 'core' arguments, or arguments for which verbs prototypically subcategorize, like actors and undergoers. The second class, represented by *sa*, mark 'oblique' arguments such as goals, beneficiaries, locations, etc. Finally, the particle *ang*, which is at the core of the theoretical controversy, must appear with one, and only one, nominal in the sentence. This nominal can fulfil any number of thematic roles: actor (185a), undergoer (185b), location (185c), instrument (185d), beneficiary (185e), etc. The thematic role of the *ang*-marked nominal coincides with a change in the morphology of the verb involving different combinations of voice (VC) and irrealis (IRR) affixes.

(185) Voice alternations in Tagalog (Foley forthcoming)

a. *b*-um-*ili ng isda sa tindahan* ang lalake
VC-buy CORE fish OBL store man
'The man bought fish in the store.'

b. *bi-bilh*-in *ng lalake sa tindahan* ang isda
IRR-buy-VC CORE man OBL store fish
'The man will buy the fish in the store.'

c. *bi-bilh*-an *ng lalake ng isda* ang tindahan
IRR-buy-VC CORE man CORE fish store
'The man will buy fish in the store.'

d. ipam-*bi-bili ng lalake ng isda* ang salapi
VC-IRR-buy CORE man CORE fish money
'The man will buy fish with the money.'

e. i-*bi-bili ng lalake ng isda* ang bata
VC-IRR-buy CORE man CORE fish child
'The man will buy fish for the child.'

Various accounts of Philippine languages mentioned above have treated *ang* as a nominative case marker signalling the subject (Bloomfield 1917, Blake 1925, Bell 1976, Kroeger 1993); absolutive case marker signalling the subject (Gerdts 1988); or topic marker (Schachter 1976, Schachter 1977, Carrier-Duncan 1985). Similarly, there have been discagreements over whether the subject is always the actor (Carrier-Duncan 1985, Gerdts 1988, Payne 1982), is the *ang*-marked constituent, or whether Tagalog has no subject at all. Most analyses have attempted to find a basic (pivot) assignment of grammatical roles and derive other sentence types from it. Foley (forthcoming) presents substantial arguments against

both the active/passive and the ergative/antipassive (as well as unergative/unaccusative) analyses, based on the lack of semantic or syntactic constraints on the combinations of nominal markers and participles and the lack of morphological distinctions on the verbs corresponding to traditional valence alternations. Instead, he suggests that Tagalog, and other Philippine languages, are of the 'symmetrical voice' type, with a system of contrasting voices where no one voice is derived from another:

> [In a symmetrical voice language,] No one NP type is preferred for PIVot choice (i.e. *ang* marked NPs in Tagalog); regardless of which choice is made, all are signalled by some overt verbal voice morphemes (e.g. *-um-*, *-in*, *-an*, *i-*, etc.) Further, other than the superposition of *ang* marking on the NP choice for PIVot, no alteration accrue to the case marking of the NPs in the clause, in marked contrast to the radical rearrangements of case marking required by marked voice options like passive or antipassive in asymmetrical languages. (Foley forthcoming)

Foley further correlates this symmetry with the fact that Philippine roots seem to be underspecified for verb vs. noun categorization, and the verbalization function falls to the 'voice' affixes like *um*, *in*, *i* etc. above. The implications of such an analysis prove to be truly constructional. Each voice in the symmetrical system is determined by verbal morphology (a combination of 'voice' and irrealis affixes) and the *ang* marking of the topic, or pivot, constituent; the assignment of core vs. non-core particles is invariant regardless of the selection of the *ang* constituent. Thus, each voice can be seen as a construction, specifying a number of discourse/information-structure (assignment of topic), semantic/syntactic (mapping between thematic role and an *ang* constituent), and morphological (verb form) properties. Similar to the middle vs. active voice constructions described in the previous section, the 'topic voice' constructions in Philippine languages are paradigmatically opposed and impose separate restraints on linguistic expressions; the different participants in these paradigmatic oppositions are not derivable from each other, but rather function as parts of a complex whole. It is to be expected that the different voice constructions may have different lexical or pragmatic properties, as well; some such differences (e.g. word order) are described for Cebuano by Payne (1994) and Sells (2001).

7.2.4 Voice in Coast Salish Languages

Descriptions of Coast Salish languages often identify sentences as being active (transitive) or passive based on morphological markers on the verbs and nominal case

marking. However, as Jelinek and Demers (1983) point out, a search for productive active/passive alternations is not always successful. In particular, both active and passive paradigms have gaps, and suppletion between them is common. Moreover, passive sentences are sometimes used where an active would be predicted semantically, and vice versa. Examples of paradigm gaps from Lummi are in (186) and (187).

(186) Lummi transitive sentences (Jelinek and Demers 1983)

x̣čí-t-oŋəs-sən 'I know you'.

x̣čí-t-sən 'I know it'.

x̣čí-t-sən cə swəyʔqəʔ 'I know the man'.

x̣čí-t-oŋəs-sxʷ 'You know me'.

x̣čí-t-sxʷ 'You know it'.

x̣čí-t-sxʷ cə swəyʔqəʔ 'You know the man'.

* ———— 'He/she knows you/me'.

x̣čí-t-s 'He/she knows it'.

x̣čí-t-s cə swəyʔqəʔ 'He/she knows the man'.

* ———— 'The man knows me/you'.

* ———— 'The man knows him'.

x̣čí-t-s cə swəyʔqəʔ cə swiʔqoʔəɬ 'The man knows the boy'.

(187) Lummi passive sentences (Jelinek and Demers 1983)

* ———— 'You are known by me'.

* ———— 'It is known by me'.

* ———— 'The man is known by me'.

* ———— 'I am known by you'.

* ———— 'It is known by you'.

* ———— 'The man is known by you'.

x̣čí-t-ŋ-sən/sxʷ 'I/you are known' (by someone).

x̣čí-t-ŋ 'It is known' (by someone.

x̣čí-t-ŋ cə swəyʔqəʔ 'The man is known' (by someone).

x̣čí-t-ŋ/sxʷ ə cə swəyʔqəʔ 'I/you are known by the man'.

x̣čí-t-ŋ ə cə swəyʔqəʔ 'It is known by the man'.

x̣čí-t-ŋ cə swiʔqoʔəɬə cə swəyʔqəʔ 'The boy is known by the man'.

As the examples show, many of the excluded sentences are those where the subject is 3rd person while the object (or oblique in the passives) is 1st or 2nd person. Jelinek and Demers (1983) propose that the paradigm gaps are motivated by an agent/animacy hierarchy in (188). In both types of sentences, the subject may not be lower on the hierarchy (further to the right) than the object/oblique. The 'excluded' sentences in Salish are similar to the types of sentences where Algonquian languages use 'inverse' marking, although relying on a slightly different agency hierarchy. (Mithun 1999, Rhodes 1990, Dodge and Leonard 2002).

(188) Salish Agent Hierarchy (based on Jelinek and Demers 1983, Dixon 1979)
1st person > 2nd person > 3rd person > Proper nouns > Human > Animate > Inanimate

The only productive active/passive alternation is possible when both arguments are of equal standing in the agent hierarchy, i.e. both 3rd person pronouns, or both humans, or both animates, etc. The use of passive sentences where active ones are expected (e.g. *you are helped by the man* instead of **the man helps you*) are attributed by Jelinek and Demers (1983) to avoidance strategies in order to conform to the agent hierarchy. However, the fact that one type of sentence is used so freely in place of another, and the lack of productive alternations puts the active/passive dichotomy into question. The agent hierarchy is clearly the guiding force in argument assignment, with the active/passive distinction being somewhat secondary.

The Coast Salish data could be interpreted fairly straightforwardly in a constructional account, with mapping constructions (similar to ones described in Chapters 3 and 4 of this thesis) as two central construction types. The process of sentence formation starting from the semantics could be described as follows. First, thematic roles (e.g. agent and patient, or actor and undergoer) are determined. Then, depending on the relative agency ranking of the two participants, grammatical relations are determined by using a mapping construction. If the agent is higher-ranked than the patient (e.g. 1st person > 3rd person), then a direct linking is applicable, assigning agent to subject and patient to object. If the patient is ranked higher, then an inverse linking applies, so that the agent is linked to object and the patient is linked to subject. The linking constructions, in turn, are associated with what Jelinek and Demers (1983) identify as active vs. passive verb morphology. The direct linking is associated with active morphology; the inverse linking, with passive

morphology. These two linking constructions may also have different implications in terms of volitionality, control, and other semantic features typically associated with active versus passive interpretations. Importantly, however, the whole process is driven by the agency ranking of the two semantic participants rather than by the active/passive comparison. If the two participants are ranked equally, the speaker makes a choice between the two linking constructions, probably guided by semantic considerations such as volitionality and control. Just as with middle voice and Philippine voice alternations, the two types of linking constructions carry a complex of formal, syntactic, and semantic properties and one cannot be said to derive from the other. At the same time, partial generalizations about the meanings of active vs. passive morphologies may arise as sub-patterns in addition to the larger constructions connected with argument linking.

7.3 Implications

Section 7.2 examined several phenomena related to voice in a diverse group of languages and suggested that in each case, what is typically identified as voice alternations is a set of oppositions between construction types, each with its own set of form and meaning properties. The differences between these construction types by and large cannot be attributed to individual morphological markers or exclusively to different argument structures. Rather, the different *combinations* of form and meaning properties distinguish the constructions from each other. Moreover, these properties can often be determined only through paradigmatic opposition with the alternative, although the opposed construction types cannot be said to derive from each other. All of these properties are naturally expressed in construction-based accounts that are ideally suited for representing complex form-meaning parings and contrasts between them. In addition, constructional approaches allow for ways to express independent or partial regularities without creating unnecessarily large sets of classifications.

Turning now to a more general question, the next section outlines some suggestions for detecting and describing constructional patterns in morphosyntax.

7.3.1 The Scope of Constructional Morphology and Morphosyntax

This dissertation has aimed to demonstrate that a Constructional Morphology framework can be used to describe all types of morphological systems, including the fully

compositional, agglutinative systems of the sort that gave rise to morpheme-based models. However, some languages have morphological systems that are particularly well-suited for constructional description, and particularly immune to morphemic accounts. Below are some of the general properties that may characterize such systems.

- Based on recurrent patterns, words can be segmented into component parts (morphs). However, it is difficult or impossible to meanings for these parts. More specifically,
  - Some morphs appear not to add any meaning at all (empty morphemes mentioned in Chapter 2).
  - Some morphs appear to mean more than one thing (portmanteau morphs)
  - Some meaning properties appear not to have corresponding morphs (zero exponents)
  - Some meaning properties appear to be expressed in a distributed fashion, where multiple properties are associated with multiple morphs, but there is no one-to-one correspondence (extended exponence)
  - Meaning differences seem to be associated with various *combinations* of morphs and can be discerned only by contrast with each other (paradigmatic effects such as Georgian screeve formation)
- Morphological, syntactic, semantic, and other properties seem to coincide in particular word or phrase types, and these word or phrase types have properties that cannot be attributed to any of their component parts (e.g. the series in Georgian).
- What seems to be a single form/distribution class has a wide variety of independent functions, but there is strong historical or other evidence against considering the different uses homonymous (e.g. Slavic reflexives, Georgian version).
- Variations in syntactic, morphological, or semantic/discourse properties are in alternations, but no one variation can be said to be 'basic', with the others derived from it (e.g. middle/active alternations).
- Some regularities are determined by combinations of, e.g. syntactic and semantic patterns, resulting in a 'mixing of levels' (e.g. Salish active/passive syntax and the agent hierarchy)
- In morphological or syntactic expression, there are partial regularities of form-meaning pairings which are determined lexically or by some other principle, but which cannot be expressed by one greater generalization. In particular, there are multiple kinds of such sub-regularities that cut across other established divisions such as declension

classes for nouns or valence classes for verbs (e.g. Russian nominal stress patterns in Chapter 2).

- There are sub-patterns that follow more general patterns but with minor modifications, and analyzing them as exceptional would miss the sub-generalization (e.g. German *bar* adjectives as described by Riehemann (2001)).

Any one of the properties above in a language should be an argument for using a constructional descriptive framework; more than one tips the balance heavily in favor of constructions. Georgian, the focus of this dissertation, could be said to exhibit all of the 'symptoms' of a constructional language, both in its morphotactics and its morphosyntax. In general, it seems that most languages with morphologically complex system have at least some non-compositional properties, and traditional descriptive grammars tend to be overwhelmingly constructional in spirit if not in form.

7.3.2 Conclusion

The goal of this dissertation has been to demonstrate the value of a construction-based approach in describing a series of complex morphological and morphosyntactic phenomena in Georgian and elsewhere. The resulting framework of Constructional Morphology draws heavily on intuitions present in traditional, structuralist grammars, but is also quite contemporary in integrating results from studies of language acquisition and processing, and from connectionist computational models. The framework was tested on several case studies of Georgian verbal morphology and morphosyntax, and the present chapter sketched out how this approach can be applied to more general morphosyntactic phenomena in other languages. A constructional view can help resolve some issues in the computational modeling of complex morphology, even within the confines of existing computational techniques, as demonstrated by the prototype model of Georgian verbs in the previous chapter.

The major intuition of Constructional Morphology, and construction-based approaches in general, is that linguistic patterns form a network of reguarlities and sub-regularities, where the most idiomatic expressions are simply at a different end in the continuum of regularity from the most general, productive patterns. Constructions, and their organization into schematic networks, is the medium in which this continuum can be expressed. Any structural pattern in language can be the locus of such patterns and sub-regularities, be it a morpheme, full word, or phrase. Further, the generalizations can involve

any combination of linguistic levels, from phonetics to pragmatics, as they are experienced simultaneously by the human cognitive system. This basic intuition can be said to apply to any language, and I hope that the framework introduced here is useful for describing the morphology of (most) any language.

Bibliography

Ackerman, Farrell, and John Moore. 1999. Syntagmatic and paradigmatic dimensions of causee encodings. *Linguistics and Philosophy* 22:1–44.

Ackerman, Farrell, and John Moore. 2001. *Proto-Properties and Grammatical Encoding. A Correspondence Theory of Argument Selection.* Stanford: CSLI.

Albright, Adam. 2002. *The identification of bases in morphological paradigms.* PhD, UCLA.

Alsina, Alex, and Sam Mchombo. 1990. The syntax of applicatives in Chichewa: Problems for a Theta Theoretic Asymmetry. *Natural Language and Linguistic Theory* 8:493–506.

Amiridze, Nino. 2006. *Reflexivization Strategies in Georgian.* Utrecht: LOT.

Anderson, Gregory D. S. 2001. Subject Version and Object Version in Tofa Auxiliary Verb Constructions. *Turkic Languages* 5(1):240–269.

Anderson, Gregory D. S. 2004. *Auxiliary Verb Constructions in Altai-Sayan Turkic.* Wiesbaden: Harrassowitz Verlag.

Anderson, Gregory D. S. to appear. Burushaski morphology. In A. Kaye (Ed.), *Morphologies of Asia and Africa.* Winona Lake: Eisenbrauns.

Anderson, Gregory D. S, and Olga Gurevich. 2005. Towards a typology of version: Formal and functional perspectives. In *Proceedings of the 41st Meeting of the Chicago Linguistic Society*, Chicago. University of Chicago Press.

Anderson, Gregory D. S., and K. David Harrison. 1999. Tyvan. *Languages of the World Materials* 257.

Anderson, Stephen R. 1984. On representations in morphology. Case, agreement and inversion in Georgian. *Natural Language and Linguistic Theory* 2:157–218.

Anderson, Stephen R. 1992. *A-Morphous Morphology*. New York: Cambridge University Press.

Aronoff, Mark. 1994. *Morphology by itself: Stems and Inflectional Classes*. Cambridge, MA: MIT Press.

Aronson, Howard I. 1982. On the status of version as a grammatical category in Georgian. *Folia Slavica* 5(1-3):66–80.

Aronson, Howard I. 1984. On Homonymy in the Georgian Verbal System. *Folia Slavica* 7(1-2):21–37.

Aronson, Howard I. 1990. *Georgian: A Reading Grammar. Corrected Edition*. Columbus, OH: Slavica Publishers.

Aronson, Howard I. 1991. Modern Georgian. In A. C. Harris (Ed.), *The Indigenous Languages of the Caucasus. Volume 1. The Kartvelian Languages*, 219–312. Delmar, NY: Caravan Books.

Aronson, Howard I. 1994. Paradigmatic and syntagmatic subject in Georgian. In H. I. Aronson (Ed.), *Non-Slavic Languages of the USSR*, 13–33. Columbus, OH: Slavica.

Aze, (F.) Richard. 1973. Clause patterns in Parengi-Gorum. In R. L. Trail (Ed.), *Patterns in Clause, Sentence, Discourse in Selected Languages of India and Nepal*, Vol. 1, 235–312. Norman, OK: SIL.

Baayen, R. Harald, James M. McQueen, Ton Dijkstra, and Robert Schreuder. 2003. Frequency effects in regular inflectional morphology: revisiting Dutch plurals. In R. H. Baayen and R. Schreuder (Eds.), *Morphological Structure in Language Processing*. Berlin: Mouton de Gruyter.

Baayen, R. Harald, and Fermín Moscoso del Prado Martín. 2005. Semantic density and past-tense formation in three Germanic languages. *Language* 81(3):666–698.

Bailey, David, Jerome A. Feldman, Srini Narayanan, and George Lakoff. 1997. Modeling embodied lexical development. In *Proceedings of the 1997 Cognitive Science Conference*, Stanford.

Baker, Mark. 1985. The mirror principle and morphosyntactic explanation. *Linguistic Inquiry* 16:373–415.

Bashir, E. 1985. Towards a semantics of the Burushaski verb. In A. Zide et al. (Eds.), *Proceeedings of the conference on participant roles: South Asia and adjacent areas*, 1–32. Bloomington: IULC.

Bates, Elizabeth, and Brian MacWhinney. 1989. Functionalism and the competition model. In B. MacWhinney and E. Bates (Eds.), *The Cross-linguistic Study of Sentence Processing*, 3–73. Cambridge: Cambridge University Press.

Beesley, Kenneth. 1996. Arabic finite-state morphological analysis and generation. In *COLING-96*, 89–94, Copenhagen.

Beesley, Kenneth, and Lauri Karttunen. 2003. *Finite-State Morphology*. New York: Cambridge University Press.

Beesley, Kenneth, and Lauri Karttunen. 2006. *Finite-State Morphology. 2nd edition*. New York: Cambridge University Press.

Bell, Sarah J. 1976. *Cebuano Subjects in Two Frameworks*. PhD thesis, UCSD.

Bell, Sarah J. 1983. Advancements and ascensions in Cebuano. In D. M. Perlmutter (Ed.), *Studies in Relational Grammar I*, 143–218. Chicago: University of Chicago Press.

Bergen, Benjamin K. 2004. The psychological reality of phonaesthemes. *Language* 80(2):290–311.

Bergen, Benjamin K., and Nancy Chang. 2005. Embodied Construction Grammar in simulation-based language understanding. In J.-O. Östman and M. Fried (Eds.), *Construction Grammar(s): Cognitive and Cross-Language Dimensions*. John Benjamins.

Bergen, Benjamin K., Nancy Chang, and Shweta Narayan. 2004. Simulated action in an Embodied Construction Grammar. In *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*, Berkeley.

Berger, H. 1998. *Die Burushaski-Sprache von Hunza und Nager. 3 vols.* Wiesbaden: Harrassowitz Verlag.

Bickel, Balthasar, and Johanna Nichols. 2006. Inflectional morphology. In T. Shopen (Ed.), *Language typology and syntactic description.* Cambridge: Cambridge University Press, 2nd edition.

Blake, Frank R. 1925. *A grammar of the Tagalog language.* New Haven: American Oriental Society.

Blevins, James P. 2001a. Paradigmatic derivation. *Transactions of the Philological Society* 99(2):211–222.

Blevins, James P (Ed.). 2001b. *Transactions of the Philological Society 2001:2. Special issue on morphological paradigms.*

Blevins, James P. 2003. Passives and impersonals. *Journal of Linguistics* 39(3):473–520.

Blevins, James P. 2004. Inflectional classes and economy. In L. Günke, G. Müller, and G. Zinfonun (Eds.), *Explorations in Nominal Inflection*, 41–85. Berlin: Mouton de Gruyter.

Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42(3).

Blevins, James P. forthcoming. *Word and Paradigm Morphology.* Stanford: CSLI.

Blevins, James P. to appear. Thematic Inversion in Georgian. *Natural Language and Linguistic Theory.*

Blevins, Juliette. 2002. Periphrastic agreement in Yurok. In F. Ackerman and J. P. Blevins (Eds.), *Proceedings of the 2002 San Diego Workshop on Periphrasis.* Stanford: CSLI.

Blevins, Juliette. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns.* Cambridge: Cambridge University Press.

Bloomfield, Leonard. 1917. *Tagalog texts with gramamtical analysis.* Vol. 3 of *University of Illinois Studies in Language and Literature.*

Bloomfield, Leonard. 1933. *Language.* Chicago: The University of Chicago Press.

Bochner, Harry. 1993. *Simplicity in Generative Morphology*. Berlin, New York: Mouton de Gruyter.

Boeder, Winfried. 1968. Über die Versionen des georgischen Verbs. *Folia Linguistica* 2:82–152.

Boeder, Winfried. 1969. Nachträge zum Aufsatz: Über die Versionen des georgischen Verbs.

Boeder, Winfried. 2005. The South Caucasian Languages. *Lingua* 115(1):5–89.

Booij, Geert. 2005. Compounding and derivation: evidence for Construction Morphology. In W. Dressler, P. Kastovsky, and F. Rainer (Eds.), *Morphology and its demarcations*, 109–132. Amsterdam: John Benjamins.

Booij, Geert. forthcoming. *Construction Morphology*.

Bošcovič, Zeljko. 2004. Clitic placement in South Slavic. *Journal of Slavic Linguistics*.

Bowerman, Melissa. 1982. Reorganizational processes in lexical and syntactic development. In E. Wanner and L. Gleitman (Eds.), *Language Acquisition: The State of the Art*, 319–346. Cambridge: Cambridge Univesity Press.

Braithwaite, Kim. 1973. *Case Shift and Verb Concord in Georgian*. PhD thesis, UT Austin.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

Bresnan, Joan, and Annie Zaenen. 1990. Deep unaccusativity in LFG. In K. Dziwirek, P. Farell, and E. Mejías-Bikandi (Eds.), *Grammatical Relations: A Cross-Theoretical Perspective*. Stanford: CSLI.

Brooks, Patricia J., and Michael Tomasello. 1999. How children constrain their argument structure constructions. *Language* 75(4):720–738.

Brooks, Patricia J., Michael Tomasello, Kelly Dodson, and Lawrence B. Lewis. 1999. Young Children's Overgeneralizations with Fixed Transitivity Verbs. *Child Development* 70(6):1325–37.

Browne, Wayles. 1993. Serbo-croat. In B. Comrie and G. G. Corbett (Eds.), *The Slavonic Languages*. London, New York: Routledge.

Bryant, John. 2003. *Constructional Analysis*. MS thesis, UC Berkeley.

Burzio, Luigi. 1996. Surface constraints versus underlying representation. In J. Durand and B. Laks (Eds.), *Current Trends in Phonology: Models and Methods*, 97–122. European Studies Research Institute, University of Salford Publications.

Burzio, Luigi. 2002. Surface-to-surface morphology: when your representations turn into constraints. In P. Boucher (Ed.), *Many Morphologies*. Cascadilla Press.

Bybee, Joan L. 1985. *Morphology: a study of the relation between meaning and form*. Amsterdam ; Philadelphia: John Benjamins.

Bybee, Joan L. 2001. *Phonology and Language Use*. Cambridge, [England] ; New York: Cambridge University Press.

Bybee, Joan L, and Paul Hopper. 2001. *Frequency and the emergence of linguistic structure*. Amsterdam ; Philadelphia: John Benjamins Pub. Co.

Cameron-Faulkner, Thea, Elena Lieven, and Michael Tomasello. 2003. A construction based analysis of child directed speech. *Cognitive Science* 27:843–873.

Carrier-Duncan, Jill. 1985. Linking of thematic roles in derivational word formation. *Linguistic Inquiry* 16(1):1–34.

Carstairs, Andrew. 1983. Paradigm economy. *Journal of Linguistics* 19(1):115–128.

Casenhiser, Devin, and Adele E. Goldberg. 2005. Fast mapping between a phrasal form and meaning. *Developmental Science* 8(6):500–508.

Cena, R.M. 1977. Patient primacy in Tagalog. In *Paper presented at the LSA Annual Meeting*, Chicago.

Chang, Nancy. 2004. A computational model of comprehension-based construction acquisition. In *Child Language Research Forum*, Stanford.

Chang, Nancy, and Olga Gurevich. 2004. Context-driven construction learning. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, 204–209, Chicago.

Cherchi, Marcello. 1997. *Modern Georgian morphosyntax: a grammatico-categorial hierarchy-based analysis with special reference to indirect verbs and passives of state.* Wiesbaden: Harrassowitz.

Chikobava, Arnold. 1984. Gruzinskij jazyk [Georgian language]. *Ezhegodnik iberijsko-kavkazskogo jazykoznanija* 11:9–35.

Chomsky, Noam. 1957. *Syntactic Structures.* The Hague: Mouton.

Chomsky, Noam. 1959. On certain formal properties of grammars. *Infromation and Control* 2:137–167.

Chomsky, Noam. 1977. *Essays on Form and Interpretation.* North-Holland: Elsevier.

Chomsky, Noam. 1991. Some notes on economy of derivation and representation. In R. Freidin (Ed.), *Principles and Parameters in Comparative Grammar. Current Studies in Linguistics*, No 20, 417–454. Cambridge, MA and London: The MIT Press.

Chomsky, Noam. 1995. *The Minimalist Program.*

Clark, Eve V. 1987. The principle of contrast: a constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition*, 1–33. Hillsdale, NJ: Lawrence Erlbaum.

Cohen-Sygal, Yael, and Shuly Wintner. 2006. Finite-state registered automata for non-concatenative morphology. *Computational Linguistics* 32(1):49–82.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective.* New York: Oxford University Press.

Croft, William, H. Bat-Zeev Shyldkrot, and Suzanne Kemmer. 1987. Diachronic semantic processes in the middle voice. In A. G. Ramat, O. Carruba, and G. Bernini (Eds.), *Papers from the VIIth International Conference on Historical Linguistics*, 179–192. Amsterdam: John Benjamins.

Culicover, Peter. 1999. *Syntactic Nuts.* Oxford: Oxford University Press.

Culicover, Peter, and Ray Jackendoff. 2005. *Simpler Syntax.* Oxford: Oxford University Press.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. Vol. 34 of *Syntax and Semantics*. New York: Academic Press.

Datukishvili, Ketevan. 1997. Some questions of computer synthesis of verb in Georgian. In R. Cooper and T. Gamkrelidze (Eds.), *Proceedings of the second Tbilisi Symposium on Language, Logic, and Computation*, 83–85, Tbilisi. Tbilisi State University.

De Guzman, Videa P. 1979. Morphological evidence for primacy of patient as subject in Tagalog. Paper presented at the LSA Annual Meeting, Los Angeles.

Demirdache, Hamida. 1997. Out of control in St'át'imets. In A. Mendikoetxea and M. Uribe-Etxebarria (Eds.), *Theoretical Issues in the Morphology-Syntax Interface, Supplement to the International Journal of Basque Linguistics and Philology*, 97–143.

Dixon, R. M. W. 1979. Ergativity. *Language* 55:59–138.

Dodge, Ellen, and Wesley Leonard. 2002. Cognitively based analysis of Miami morphology. Manuscript, UC Berkeley.

Dodge, Ellen, and Abby C Wright. 2002. Herds of wildebeest, flasks of vodka, heaps of trouble: An Embodied Construction Grammar approach to English measure phrases. In J. Larson and M. Paster (Eds.), *Proceedings of the 28th Annual Meeting of the Berkeley Lingusitics Society*, 75–86, Berkeley.

Dodson, Kelly, and Michael Tomasello. 1998. Acquiring the Transitive Construction in English: The Role of Animacy and Pronouns. *Journal of Child Language* 25(3):605–22.

Doerfer, G. 1988. *Das Chorasantürkische*. Wiesbaden: Harrassowitz Verlag.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3):547–619.

Dąbrowska, Ewa. 2002. Rules or Schemas? Evidence from Polish.

Erelt, Tiiu (Ed.). 1999. *Eesti keele sõnaraamat*. Tallinn: Eesti Keele Sihtasutus.

Feldman, Jerome A. 1988. Connectionist representation of concepts. In D. Waltz and J. A. Feldman (Eds.), *Connectionist Models and Their Applications*. Ablex Publishing Company.

Feldman, Jerome A. 2006. *From Molecule to Metaphor: A Neural Theory of Language*. Cambridge, MA: MIT Press.

Fillmore, Charles J. 1968. The case for Case. In E. Bach and R. T. Harms (Eds.), *Universals in Linguistic Theory*, 1–88. Holt, Rinehart and Winston.

Fillmore, Charles J. 1982. Frame semantics. In *Linguistics in the Morning Calm*, 111–137. Seoul: Hanshin Publishing Co.

Fillmore, Charles J. 1988. The Mechanisms of "Construction Grammar". *BLS* 14:35–55.

Fillmore, Charles J., Paul Kay, Laura Michaelis, and Ivan Sag. forthcoming. *Construction Grammar*. Stanford: CSLI.

Fillmore, Charles J., Paul Kay, and Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of Let alone. *Language* 64(3):501–538.

Flickinger, Daniel, Carl Pollard, and Thomas Wasow. 1985. Structure sharing in lexical representation. In *Proceedings of the Twenty-third Annual Meeting of the Association for Computational Linguistics*, 262–267, Chicago.

Foley, William A. forthcoming. Symmetrical voice systems and precategoriality in Philippine languages. Manuscript.

Franks, Stephen, and Tracy Holloway King. 2000. *A Handbook of Slavic Clitics*. Oxford: Oxford University Press.

Fried, Mirjam. 1994. Second-position clitics in Czech: Syntactic or phonological? *Lingua* 94:155–175.

Garrett, Andrew. forthcoming. Paradigmatic uniformity and markedness. In J. Good (Ed.), *Explaining Linguistic Universals: Historical Convergence and Universal Grammar*. Oxford: Oxford University Press.

Garrett, Andrew, and Juliette Blevins. forthcoming. Analogical morphophonology. In K. Hanson and S. Inkelas (Eds.), *The nature of the word: Essays in honor of Paul Kiparsky*. Cambridge, Mass.: MIT Press.

Gasser, Michael, and Eliana Colunga. 2002. Pattern learning in infants and neural networks. In P. Quinlan (Ed.), *Connectionist Models of Development*. Brighton: Psychology Press.

Geniušiene, Emma. 1987. *The Typology of Reflexives*. Berlin: Mouton de Gruyter.

Gentner, Dedre, and Arthur B. Markman. 1997. Structure Mapping in Analogy and Similarity. *American Psychologist* 52(1):45–56.

Gentner, Dedre, and José Medina. 1998. Similarity and the development of rules. *Cognition* 65:263–297.

Gerdts, Donna B. 1988. Antipassive and causatives in Ilokano: evidence for an ergative analysis. In P. McGinn (Ed.), *Studies in Austronesian Linguistics*, 295–321. Athens, OH: Ohio University Press.

Gerritsen, Nelleke. 1990. *Russian Reflexive Verbs: In search of unity and diversity*. Vol. 15 of *Studies in Slavic and General Linguistics*.

Ginzburg, Johathan, and Ivan A. Sag. 2000. *Interrogative Investigations: the form, meaning and use of English interrogative constructions*. Stanford: CSLI.

Givón, Talmy. 1979. *On Understanding Grammar*. New York: Academic Press.

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*.

Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Goldberg, Adele E., Devin Casenhiser, and Nitya Sethuraman. 2004. Learning argument structure generalizations 14. *Cognitive Linguistics* 14:289–316.

Goletiani, G.G. 1970. *Sopostavitelnaja grammatika russkogo i gruzinskogo jazykov. Ch. I. Morfologija. [Comparative grammar of Russian and Georgian languages. Part I. Morphology]*. Tbilisi: Izdatel'stvo tbilisskogo universiteta.

Gurevich, Olga. 2005a. Computing non-concatenative morphology: The case of Georgian. In *To appear in Proceedings of the Conference on Lesser-Used Languages and Computational Linguistics*, Bolzano, Italy.

Gurevich, Olga. 2005b. *Steal me an apple*: Version in Georgian. In *To appear in Proceedings of the 9th Texas Linguistic Society Conference*, Austin, TX. CSLI.

Gurevich, Olga. 2006. A finite-state model of Georgian verbal morphology. In *HLT/NAACL 2006*, New York.

Hahn, R. [w/ Ablahat Ibrahim]. 1991. *Spoken Uyghur*. Seattle: University of Washington Press.

Halle, Morris, and Alec Marantz. 1994. Distributed Morphology and the pieces of inflection. In K. Hale and S. J. Keyser (Eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*. MIT Press.

Harris, Alice C. 1981. *Georgian Syntax: A Study in Relational Grammar*. Cambridge: Cambridge University Press.

Harris, Alice C. 1985. *Diachronic Syntax: The Kartvelian Case*. Orlando: Academic Press.

Harris, Alice C. 1989. Reply to B. George Hewitt. *Revue des études géorgiennes et caucasiennes* 5:203–225.

Harris, Alice C. 1990. Georgian: a language with active case marking. *Lingua* 80:47–65.

Harris, Alice C. 2004. History in Support of Synchrony. *Proceedings of the 30th Annual Meeting of the Berkeley Linguistic Society*.

Harris, Zellig S. 1942. Morpheme Alternants in Linguistic Analysis. *Language* 18:169–180.

Harris, Zellig S. 1951. *Methods in Structural Linguistics*. Chicago: University of Chicago Press.

Haspelmath, Martin. 2002. *Understanding Morphology*. New York: Oxford University Press.

Hay, Jennifer B. 2001. Lexical frequency in morphology: is everything relative? *Linguistics* 39(6):1041–1070.

Hay, Jennifer B., and R. Harald Baayen. 2005. Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences* 9(7):342–348.

Hewitt, B. G. 1987a. Georgian: ergative or active? *Lingua* 71:319–340.

Hewitt, B. G. 1987b. Review of 'Diachronic Syntax: the Kartvelian case by Alice C. Harris'. *Revue des études géorgiennes et caucasiennes* 3:173–213.

Hewitt, B. G. 1995a. Georgian — ergative, active, or what? In D. Bennett, T. Bynon, and B. G. Hewitt (Eds.), *Subject, Voice, and Ergativity. Selected Essays*, 202–217. London: School of Oriental and African Studies, University of London.

Hewitt, B. G. 1995b. *Georgian: A Structural Reference Grammar*. Amsterdam: John Benjamins.

Hillery, PJ. 2001. The morphology of the Georgian Aorist. available from http://www.-armazi.com.

Hippisley, Andrew. 1999. Word formation rules in a default inheritance framework: a Network Morphology account of Russian personal nouns. In G. Booij and J. van Marle (Eds.), *Yearbook of Morphology 1999*, 221–261. Kluwer Academic Publishers.

Hockett, Charles F. 1947. Problems of Morphemic Analysis. *Language* 23:321–43.

Hockett, Charles F. 1954. Two Models of Grammatical Description. *Word* 10:210–31.

Hoff, Erika. 2001. *Language Development*. Belmont, CA: Wadsworth. Second edition edition.

Holisky, Dee Ann. 1981a. *Aspect and Georgian medial verbs*. Delmar, N.Y.: Caravan Books.

Holisky, Dee Ann. 1981b. Aspect theory and Georgian aspect. In P. J. Tedeschi and A. Zaenen (Eds.), *Syntax and Semantics: Tense and Aspect*, Vol. Volume 14, 127–144. New York: Academic Press.

Hyman, Larry M. 2002. Suffix ordering in Bantu: A morpho-centric approach. In G. Booij and J. van Marle (Eds.), *Yearbook of Morphology*, 245–281.

Imedadze, Natela, and Kevin Tuite. 1992. The acquisition of Georgian. In D. I. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition*, Vol. 3, 39–109. London: LEA.

Imnaishvili, Ivane. 1968. vnebiti gvaris zmnata taviseburebani dzvel kartulshi [characteristics of passive-voice verbs in old georgian]. *Dzveli kartuli enis k'atedris shromebi* 11:27–54.

Inkelas, Sharon, and Cheryl Zoll. 2005. *A Morphological Theory of Reduplication*.

Israeli, Alina. 1997. *Semantics and Pragmatics of the "Reflexive" Verbs in Russian*. München: Verlag Otto Sagner.

Jackendoff, Ray. 1975. Morphological and Semantic Regularities in the Lexicon. *Language* 51:639–671.

Janda, Laura A., and Stephen J. Clancy. 2002. *The Case Book for Russian*. Bloomington, IN: Slavica.

Jelinek, Eloise, and Richard A. Demers. 1983. The agent hierarchy and voice in some coast salish languages. *International Journal of American Linguistics* 49(2):167–185.

Johnson, Keith. 2005a. Decisions and mechanisms in exemplar-based phonology. In *UC Berkeley Phonology Lab Annual Report 2005*, 289–311.

Johnson, Keith. 2005b. Resonance in an exemplar-based lexicon: the emergence of social identity and phonology. In *UC Berkeley Phonology Lab Annual Report 2005*, 95–128.

Jorbenadze, Bessarion. 1983. *zmnis xmovanp'repiksuli c'armoeba kartulshi [Vowel-prefix derivation of the verb in Georgian]*. Tbilisi: Tbilisi University Press.

Jurafsky, Daniel, and James H. Martin. 2000. *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, NJ: Prentice Hall.

Kaplan, Ronald M., and Joan Bresnan. 1982. Lexical-functional grammar: A formal system for grammatical representation. In J. Bresnan (Ed.), *The Mental Representation of Grammatical Relations*, 173–281. Cambridge, MA: MIT Press.

Kaplan, Ronald M., and Martin Kay. 1981. Phonological rules and finite-state transducers. abstract. In *Linguistic Society of American Meeting Handbook, Fifty-Sixth Annual Meeting*, New York.

Kaplan, Ronald M., and Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics* 20(3):331–378.

Karttunen, Lauri. 1983. KIMMO: a general morphological processor. In M. Dalrymple, E. Doron, J. Goggin, B. Goodman, and J. McCarthy (Eds.), *Texas Linguistic Forum, Vol. 22*, 165–186. Austin, TX: Department of Linguistics, The University of Texas at Austin.

Kay, Paul. 2002. An informal sketch of a formal architecture for Construction Grammar. *Grammars* 5(1):1–19.

Kay, Paul, and Charles J. Fillmore. 1999. Grammatical Constructions and Linguistic Generalizations: The What's X Doing Y? Construction. *Language* 75(1):1–33.

Kemmer, Suzanne. 1994. Middle voice, transitivity, and the elaboration of events. In B. Fox and P. J. Hopper (Eds.), *Voice: Form and Function*, 179–230. Amsterdam: John Benjamins.

King, Tracy Holloway. 1994. Agentivity and the Georgian Ergative. In S. Gahl and A. Dolbey (Eds.), *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society February 18-21, 1994: General Session: Dedicated to the Contributions of Charles J Fillmore.*

Klaiman, M.H. 1991. *Grammatical Voice.* Cambridge: Cambridge University Press.

Koenig, Jean-Pierre. 1999. *Lexical Relations.* Stanford: CSLI.

Korkina, E. I. et al. 1982. *Grammatika sovremennogo jakutskogo literaturnogo jazyka [Grammar of modern literary Yakut].* Moscow: Nauka.

Koskenniemi, Kimmo. 1983. Two-level morphology: A general computational model for word-form recognition and production. Technical report, University of Helsinki, Department of General Linguistics.

Kroeger, Paul. 1993. *Phrase Structure and Grammatical Relations in Tagalog.* Stanford: CSLI.

Krott, Andrea, R. Harald Baayen, and Robert Schreuder. 2001. Analogy in morphology: modeling the choice of linking morphemes in Dutch. *Linguistics* 39(1):51–93.

Krott, Andrea, Robert Schreuder, and R. Harald Baayen. 2002. Analogical hierarchy: exemplar-based modeling of linkers in Dutch noun-noun compounds. In R. Skousen, D. Lonsdale, and D. B. Parkinson (Eds.), *Analogical modeling: an exemplar-based approach to language*, 181–206. Amsterdam: John Benjamis.

Krott, Andrea, Robert Schreuder, R. Harald Baayen, and Wolfgang U. Dressler. forthcoming. Analogical effects on linking elements in German. *Language and Cognitive Processes*.

Kuczaj, Stan A. 1977. The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior* 16:589–600.

Kurtsikidze, Shorena. forthcoming. *Essentials of Georgian Grammar*. Lincom.

Kuryłowicz, Jerzy. 1949a. La nature des procs dits "analogiques". *Acta Linguistica* 5:15–37.

Kuryłowicz, Jerzy. 1949b. Linguistique et théorie du signe. *Journal de Psychologie* 170–80.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind.* Chicago: University of Chicago Press.

Lakoff, George, and Mark Johnson. 1980. *Metaphors we Live by.* Chicago: University of Chicago Press.

Langacker, Ronald W. 1987a. *Foundations of Cognitive Grammar.* Vol. 1. Stanford: Stanford University Press.

Langacker, Ronald W. 1987b. *Foundations of Cognitive Grammar.* Vol. 2. Stanford: Stanford University Press.

Langacker, Ronald W. 1991. *Concept, Image, and Symbol. The Cognitive Basis of Grammar.* New York: Mouton de Gruyter.

Legendre, G., J. Grimshaw, and S. Vikner (Eds.). 2001. *Optimality-Theoretic Syntax.* Cambridge, MA: MIT Press.

Lieber, Rochelle. 1980. *On the Organization of the Lexicon.* PhD dissertation, MIT.

Lieven, Elena, Heike Behrens, Jennifer Speares, and Michael Tomasello. 2003. Early syntactic creativity: a usage-based approach. *Journal of Child Language* 30(2):333–370.

Loewenstein, Jeffrey, and Dedre Gentner. 2005. Relational language and the development of relational mapping. *Cognitive Psychology* 50:315–353.

Lyons, John. 1968. *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.

Mach'avariani, Givi. 1988. k'auzat'ivis k'at'eboria kartvelur enebshi [the category of causative in Kartvelian languages]. *Iberiul-k'avk'asiuri enatnecniereba [Iberio-Caucasian Linguistics]* 27:62–106.

Mach'avariani, Maya. 1987. *Kcevis gramatik'uli k'at'egoriis semant'ik'a [Semantics of the grammatical category of version]*. Tbilisi: Mecniereba.

Manning, Christopher D., and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: The MIT Press.

Margvelani, L., L. Samsonadze, and N Javashvili. 1997. Computer aid for Georgian morphology teaching. In R. Cooper and T. Gamkrelidze (Eds.), *Proceedings of the second Tbilisi Symposium on Language, Logic, and Computation*, 208–217, Tbilisi. Tbilisi State University.

Matthews, P. H. 1972. *Inflectional morphology: a heoretical study based on aspects of Latin verb conjugation*. Cambridge: Cambridge University Press.

Matthews, P. H. 1991. *Morphology*. New York: Cambridge University Press. 2nd edition.

Matthews, Peter H. 1999. Zellig Sabbettai Harris. *Language* 75(1):112–119.

McCarthy, John. 2005. Optimal paradigms. In L. Downing, T. A. Hall, and R. Raffelsiefen (Eds.), *Paradigms in Phonological Theory*. Oxford: Oxford University Press.

McClelland, J., D. Rumelhart, and the PDP Research Group (Eds.). 1986. *Parallel Distributed processing: Exploration in the microstructure of cognition.* Cambridge, MA: MIT Press.

Melikishvili, Damana. 2001. *Kartuli zmnis ughlebis sist'ema [Conjugation system of the Georgian verb]*. Tbilisi: Logos presi.

Merlan, Francesca. 1982. Another look at Georgian "inversion". *Folia Slavica* 5(1-3):294–312.

Meurer, Paul. 1999. A morphological parser for Georgian. In *Batumi Symposium on Language, Logic and Computation*, Batumi.

Michaelis, Laura, and Knud Lambrecht. 1996. Toward a construction-based theory of language function: The case of nominal extraposition. *Language* 72(1):215–247.

Michaelis, Laura, and Josef Ruppenhofer. 2001. *Beyond Alternations. A Constructional Model of the German Applicative Pattern*. Stanford: CSLI.

Mithun, Marianne. 1994. The implications of ergativity for a Philippine voice system. In B. Fox and P. J. Hopper (Eds.), *Voice. Form and Function*, 247–278. Amsterdam: John Benjamins.

Mithun, Marianne. 1999. *The Languages of Native North America*. Cambridge: Cambridge University Press.

Morpurgo Davies, Anna. 1978. Analogy, segmentation and the early neogrammarians. *Transactions of the Philological Society* 76:36–60.

Morpurgo Davies, Anna. 1992. *Nineteenth-century linguistics*. London: Longman.

Moscoso del Prado Martín, Fermín, Avital Deutch, Ram Frost, Robert Schreuder, Nivja H. de Jong, and R. Harald Baayen. 2005. Changing places: A cross-language perspective on frequency and family size in dutch and hebrew. *Journal of Memory and Language* 53(4):496–512.

Narayan, Shweta, Benjamin K. Bergen, and Zachary Weinberg. 2004. Embodied verbal semantics: Evidence from a lexical matching task. In *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society*, Berkeley.

Narayanan, Srini. 1999. Moving right along: A computational model of metaphoric reasoning about events. In *Proceedings of the National Conference on Artificial Intelligence (AAAI '99)*, 121–128, Orlando, FL. AAAI Press.

Nozadze, L. 1974. medioakt'iv zmnata c'armoebis zogi sak'itxi kartulši (c'inasc'ar moxseneba) [Some issues concerning the formation of medioactive verbs in Georgian (preliminary report)]. *Iberiul-k'avk'asiuri enatnecniereba [Iberio-Caucasian Linguistics]* 19:25–53.

Oflazer, Kemal, and Sharon Inkelas. 2006. The architecture and the implementation of a finite state pronunciation lexicon for Turkish. *Computer Speech and Language* 20(1):80–106.

Orgun, Cemil Orhan. 1996. *Sign-Based Morphology and Phonology.* PhD thesis, UC Berkeley.

Orgun, Cemil Orhan. 1997. Cyclic and noncyclic effects in a declarative grammar. In G. Booij and J. van Marle (Eds.), *Yearbook of Morphology 1997*, 179–218.

Orgun, Cemil Orhan. 1999. Sign-Based Morphology: a declarative theory of phonology-morphology interleaving. In B. Hermans and M. van Oostendorp (Eds.), *The derivational residue in phonological Optimality Theory.* Amsterdam: John Benjamins.

Orgun, Cemil Orhan, and Sharon Inkelas. 2001. Reconsidering bracket erasure. In G. Booij and J. van Marle (Eds.), *Yearbook of Morphology 2001*, 115–46.

Payne, Thomas E. 1982. Role and reference related subject properties and ergativity in Yup'ik Eskimo and Tagalog. *Studies in Language* 6(1):75–106.

Payne, Thomas E. 1994. The pragmatics of voice in a Philippine language: Actor-focus and goal-focus in Cebuano narrative. In T. Givón (Ed.), *Voice and Inversion*, 317–364. Amsterdam: John Benjamins.

Perlmutter, David M. 1983. Towards a universal characterization of passivization. In D. M. Perlmutter (Ed.), *Studies in Relational Grammar I*, 3–29. Chicago: University of Chicago Press.

Perrott, David, Dedre Gentner, and Galen V. Bodenhausen. to appear. Resistance is futile: The unwitting insertion of analogical inferences in memory. *Psychonomic Bulletin and Review.*

Peterson, David A. 1999. *Discourse-functional, historical, and typological aspects of applicative constructions.* PhD thesis, UC Berkeley.

Peterson, David A. forthcoming. *Applicatives.*

Pinker, Stephen. 1989. *Learnability and Cognition: The acquisition of verb-argument structure.* Cambridge, MA: Harvard University Press.

Plaut, David C., and Laura M. Gonnerman. 2000. Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes* 15(4-5):445–485.

Prince, Alan, and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. rutgers optimality theory archive.

Ramos, Teresita V. 1974. *The Case System of Tagalog Verbs*. Pacific Linguistics B-27.

Rassadin, V. I. 1994. *Törœæn soot [Native word]*. Irkutsk: Vostochno-sibirskoe knizhnoe izdatel'stvo.

Reddy, Michael. 1979. The Conduit metaphor. In A. Ortony (Ed.), *Metaphor and Thought*, 284–324. Cambridge: Cambridge University Press.

Regier, Terry. 1996. *The Human Semantic Potential*. Cambridge, MA: MIT Press.

Regier, Terry, Bryce Corrigan, Rachael Cabasaan, Amanda Woodward, Michael Gasser, and Linda Smith. 2001. The Emergence of Words. *Proceedings of the Annual Conference of the Cognitive Science Society* 23:Online.

Rhodes, Richard A. 1990. Obviation, inversion, and topic rank in Ojibwa. In D. J. Costa (Ed.), *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society: Special Section on General Topics in American Indian Linguistics*, 101–115, Berkeley.

Riehemann, Susanne Z. 2001. *A Constructional Approach to Idioms and Word Formation*. PhD thesis, Stanford.

Robins, R.H. 1959. In defense of WP. *Transactions of the Philological Society* 116–144.

Sag, Ivan, Thomas Wasow, and Emily Bender. 2003. *Syntactic theory: a formal introduction*. Stanford: CSLI.

Sarjveladze, Zurab. 1987. dep'onensu zmnata shesaxeb dzvel kartulshi [about deponent verbs in old georgian]. In E. Khintibidze (Ed.), *Ak'ak'i Shanidze — 100: saiubileo k'rebuli*, 102–105. Tbilisi: Tbilisi University Press.

Savage, Ceri, Elena Lieven, Anna Theakston, and Michael Tomasello. 2003. Testing the abstractness of children's linguistic representations: lexical and structural priming of syntactic constructions in young children. *Developmental Science* 6(5):557–567.

Schachter, Paul. 1976. The subject in Philippine languages: topic, actor, actor-topic or none of the above. In C. Li (Ed.), *Subject and Topic*, 491–518. New York: Academic Press.

Schachter, Paul. 1977. Reference-related and role-related properties of subjects. In *Syntax and Semantics 8*, 279–306. New York: Academic Press.

Schenker, Alexander M. 1995. *The Dawn of Slavic: An Introduction to Slavic Philology.* New Haven and London: Yale University Press.

Schmidt, Karl Horst. 1962. Zum Passivum in Georgischen und in indogermanischen Sprachen. *Bedi Kartlisa* 13-14:116–126.

Schmidt, Karl Horst. 1965. Indogermanisches Medium und Sataviso im Georgischen. *Bedi Kartlisa* 19-20:129–135.

Selkirk, Elisabeth. 1982. *The Syntax of Words.* Cambridge, MA: MIT Press.

Sells, Peter. 2001. Form and function in the typology of grammatical voice systems. In G. Legendre, J. Grimshaw, and S. Vikner (Eds.), *Optimality-Theoretic Syntax*, 356–391. Cambridge, MA: MIT Press.

Shanidze, Akaki. 1926. Kartuli zmnis sarcevi [with a resume in French]. *T'bilisis Universit'et'is Moambe [Tbilisi University Bulletin]* 6:312–338.

Shanidze, Akaki. 1953. *Kartuli enis gramat'ik'is sapudzvlebi I. Morpologia. [Foundations of the grammar of the Georgian language I. Morphology.].* Tbilisi: Tbilisi University gamomcebloba [publishing house].

Shanidze, Akaki. 1980. *Kartuli enis gramat'ik'is sapudzvlebi I. Morpologia. [Foundations of the grammar of the Georgian language I. Morphology.].* Tbilisi: Tbilisi University gamomcebloba [publishing house].

Shibatani, Masayoshi. 1988. Voice in Philippine languages. In M. Shibatani (Ed.), *Passive and Voice*, 85–142. Amsterdam: John Benjamins.

Siewierska, Anna. 1988. The passive in Slavic. In M. Shibatani (Ed.), *Passive and Voice*, 243–290. Amsterdam: John Benjamins.

Skousen, Royal, Deryle Londsdale, Dilworth B. Parkinson, and Conference on Analogical Modeling of the Language. 2002. *Analogical modeling: an exemplar-based approach to language*. Amsterdam: John Benjamins.

Skvorcov, M. I. 1999. *Chavashla-vyrasla tata vyrasla-chavashla slovar' [Chuvash-Russian and Russian-Chuvash Dictionary]*. Cheboksary: Chuvknigizdat.

Slobin, Dan I. 1985. Crosslinguistic evidence for the language-making capacity. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition*, Vol. 2, 1157–1257. Hillsdale, N.J.: Lawrence Erlbaum.

Steele, Susan. 1995. Towards a theory of morphological information. *Language* 71(2):260–309.

Steele, Susan. 2002. Many plurals: Inflection, informational additivity, and morphological processes. In P. Boucher and M. Plenat (Eds.), *Many Morphologies*. Cascadilla.

Stump, Gregory T. 1993. On rules of referral. *Language* 69(3):449–479.

Stump, Gregory T. 2001. *Inflectional morphology: a theory of paradigm structure*. Cambridge ; New York: Cambridge University Press.

Sweetser, Eve. 1990. *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge University Press.

Tandashvili, M. 1999. *Main Principles of Computer-Aided Modeling, http://titus.uni-frankfurt.de/personal/manana/refeng.htm*. Tbilisi Habilitation.

Timberlake, Alan. 2004. *A Reference Grammar of Russian*. Cambridge University Press.

Tomasello, Michael. 2000. The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences* 4:156–163.

Tomasello, Michael. 2003. *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge, Mass.: Harvard University Press.

Tomasello, Michael. 2006. Acquiring linguistic constructions. In D. Kuhn and R. Siegler (Eds.), *Handbook of Child Psychology*, Vol. 1, 1–48. Wiley.

Tschenkéli, Kita. 1958. *Einführung in die georgische Sprache*. Zurich: Amirami Verlag.

Tschenkéli, Kita. 1960-1974. *Georgisch-Deutsches Worterbuch*. Zurich: Amirani-Verlag.

Tuite, Kevin. 1994. Syntactic subject in Georgian. In H. I. Aronson (Ed.), *Non-Slavic Languages of the USSR.*, 218–227. Columbus, OH: Slavica Publishers.

Tuite, Kevin. 1997. *Svan*. Munich: Lincom Europa.

Tuite, Kevin. 2002. Deponent verbs in Georgian. In W. Bublitz, M. von Roncador, and H. Vater (Eds.), *Philology, Typology and Language Structure. Festschrift for Winfried Boeder on the Occasion of his 65th Birthday*, 375–389. Frankfurt am Main: Peter Lang.

Tuite, Kevin. 2003. Liminal morphosyntax: Georgian deponents and their kin. In *Proceedings of the 39th Chicago Linguistic Society Conference*, Chicago.

Vendler, Zeno. 1967. Verbs and times. In Z. Vendler (Ed.), *Linguistics in Philosophy*, 97–121. Ithaca: Cornell University Press.

Vinogradov, V.V. 1947. *Russkij Jazyk (Grammatičeskoe učenie o slove)*. Moskva: Vysšaja škola.

Vogt, Hans. 1938. Esquisse d'une grammaire du Géorgien moderne. *Norsk Tidsskrift for Sprogvidenskab* 10:5–188.

von Gabain, A. 1974. *Alttürkische Grammatik*. Wiesbaden: Harrassowitz.

Šarič, Ljiljana. 2003. On the semantics of the "dative of possession" in the Slavic languages: An analysis on the basis of Russian, Polish, Croatian/Serbian and Slovenian examples. *Glossos* 3:Online.

Švedova, N et al. 1980. *Russkaja grammatika*. Vol. 1. Moskva: Nauka.

Wedel, Andrew. 2003. *The self-organization of sound-systems: Spontaneous development of natural phonology in an analogical lexicon.* PhD thesis, UC Santa Cruz.

Woodcock, E.C. 1959. *A new Latin syntax*. Cambridge, MA: Harvard University Press.

Wright, Abby C, and Andreas Kathol. 2003. When a head is not a head: A constructional approach to exocentricity in English. In J.-B. Kim and S. Wechsler (Eds.), *Proceedings of the 9th International Conference on Head-Driven Phrase Structure Grammar*, 373–389, Stanford. CSLI.

Wunderlich, Dieter. 1991. How do prepositional phrases fit into compositional syntax and semantics? *Linguistics* 29:591–621.

Yakubovich, Ilya. 2006. Prehistoric contacts between Hittite and Luvian: The case of reflexive pronouns. manuscript.

Zwicky, Arnold M. 1985. How to describe inflection. In *Proceedings of the Annual Meeting of the Berkeley Linguistic Society*, Vol. 11, 372–386.

Zwicky, Arnold M. 2001. Radical constructionism. Paper given at the 1st International Conference on Construction Grammar. berkeley.

Zwitserlood, Pienie. 1996. Form priming. *Language and Cognitive Processes* 11:589–96.