

UC Irvine

UC Irvine Previously Published Works

Title

Content analysis, cultural grammars, and computers

Permalink

<https://escholarship.org/uc/item/7b7021bq>

Journal

Qualitative Sociology, 14(4)

ISSN

0162-0436

Authors

Colby, Benjamin N

Kennedy, Sayuri

Milanesi, Louis

Publication Date

1991-12-01

DOI

10.1007/bf00989646

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at

<https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Content Analysis, Cultural Grammars, and Computers

Benjamin N. Colby, Sayuri Kennedy, and Louis Milanese

This paper describes current theoretical work in a cognitive anthropological approach to the analysis of oral literature. A theory of narrative grammar is presented along with the description of a computer program, SAGE, that is currently under development. SAGE facilitates the labelling of clauses and sentences by different types of semantic features. It maps and counts the occurrences of these in the stories analyzed.

Specialized areas of culture can be described in terms of grammars or compositional rules. One can, for example, derive a cultural grammar from a particular expressive form—an art form or narrative genre—if it is within a small enough community of individuals who share the same language and cultural presuppositions. These grammars have the potential for providing insight into processes of the larger culture system. A narrative grammar, for instance, can yield information about cultural dynamics on a broad scale because it constitutes both a cultural logic and a vast set of expectations about objects and events in the world. These objects and events are linked to motivations, goals and values that are common to the social group that uses the narrative. In addition to narrative are other cultural forms, all communicative in nature, which have been described by grammars or rules: body painting (Faris 1972), writing systems (Watt 1986), folk tales (Colby 1973), divination (Colby and Colby 1981) and classroom interaction routines (Mehan 1979).

Of the many possibilities for grammar-like descriptions the most interesting are to be found in narrative grammars of plot structure. The Rus-

Address correspondence to: B. N. Colby, School of Social Sciences, University of California at Irvine, Irvine, CA 92717.

sian folklorist, Vladimir Propp (1968) was one of the first to show a set of plot units to be operating in a regular fashion. In spite of Propp's work, formal analysis and testing have not been emphasized in structural studies of myth and folktales. Usually narrative structures are derived intuitively and announced by fiat, rather than being shown to have some valid distributional basis as Propp did. The plot units Propp identified, however, tell us something about how reality and fantasy were constructed in terms of a narrative and behavioral-situational logic which is expressed in part by a normative sequence, which, with minor variations, accounted for a sample of 100 Russian fairy tales. Propp himself thought this stereotyped sequence was universal. Subsequent analyses of other folktales from other cultural groups suggest that it was specific to time, place and genre (Colby 1973). Propp's units, then, constitute a system that can account for other Russian fairy tales but not for fairy tales from other culture areas.

The phenomena Propp analyzed, while not described as a grammar, can be reanalyzed and described in terms of grammar-like rules. Such a grammar can account for all the folktales or stories (as specified by the grammar) during some period in some locality of culture users, as indeed the data seem to do in Propp's corpus. The key process in writing a grammar, therefore, is in the testing of the grammar with new narrative productions in the same genre. That is, some areas, such as Ireland, have several distinct types of folktales. Each type can be described as a particular genre and would, presumably require a slightly different grammar to cover all of the examples of that genre. Testing is critical for the validation of cultural grammars of any type, whether they be trickster stories, hero tales or some other type.

EIDOCHRONIC ANALYSIS

We are using the term, eidon, for the basic unit in narrative grammar. In earlier work (Colby 1973) an eidon was described as an "eidochronic unit," a unit in time or sequence such as exists in a narrative. Since that time we have come to use the term more broadly to cover other areas of what Bateson characterized as the "eidos" of a cultural system (Bateson 1936). In this wider application eidon can be conceptualized as any cognitive image or concept which exists in a postulated cognitive system that has cultural reality—that can be readily communicated among the members of some culturally defined social group. The former, narrower meaning can then be expressed as "plot eidon," a member of a set of narrative events and circumstances that form the basic plot constituents of well-formed narratives of a particular genre and cultural system and that follow a series

of rules that govern their sequence. To show how this works, part of the synopsis for an Eskimo folktale, *The Headband* (Spencer 1959: 388-390) is presented below in terms of eidons:

The headband

Initial Situation (IS): an orphan and his grandmother live by the sea. The orphan is mistreated by the other boys but is protected by a rich man and his son. [The protagonist is introduced and located in space and social relationships:]

Villainy (VL): Some hunters of the village, including the rich man's son, do not return from a hunting trip. [It is understood, and later stated, that an adversary has acted against members of the protagonist's group:]

Highlighting of Facilitation: The orphan asks the rich man to outfit him for a search, but the rich man refuses, saying he does not want him to go on such a dangerous trip. [The coming actions of the protagonist are highlighted by the emphasis on the danger involved, which is the reason for the refusal of his request:]

Facilitation (Fc): The boy is given a magical headband, some arrows and other items. [The protagonist is outfitted for the coming engagement:]

Departure (DP): The boy leaves

Etc.

In the remainder of the story the boy approaches a village where he is attacked by an ogre. He kills the ogre with the magic arrows but the people of the village turn out to be magical animals and they pursue him. He uses the headband to make himself invisible to the pursuers. He escapes and returns triumphantly to his village.

The full story can be coded in terms of the following eidon string:

IS (Initial Situation) VI (Villainy) Fc (Facilitation) Dp (Departure) Pc (Perception) Hs (Hospitality) Ak (Attack) St. (Struggle) Es (Escape) El (Elimination) Dc (Deception) Vc (Victory) Es (Escape) Oe (Obstacle or Evasion) Rt (Return) Gr (Group of Reference) Se (Settlement) Co (Narrator's Comment) Oe (Obstacle or Evasion) At (Attainment)

Each eidon in the above example is further particularized by subscripted numbers (not given here) to indicate which of a series of eidon varieties for each eidon type is represented in the story. For example, one Eskimo plot eidon, "magical Engagement (Me)" has two varieties: Me1 "The protagonist engages in a magical contest" and Me2, "The protagonist engages in a unilateral magical action against the adversary."

At the most general level of analysis stories may be initially conceptualized simply as having beginnings, middles and ends. However when one makes distinctions between what are commonly described as stories, a series of events, or actions with some point or purpose behind the telling, the difference between these three sections goes beyond their positions to their content, i.e., to what might be called psychological function. Thus when beginning an analysis of a well-told story the analyst can think of it in terms of psychological functions—as having a motivational section in which some event creates a problem or difficult situation that gets the story moving, an action or engagement section in which attempts are made to

cope with the difficulty or remove it, and a resolution section where the situation has been returned to normal, usually with some added benefit over the initial situation when the story began. A distributional analysis (i.e., a study of how elements are distributed in a story—what kinds of environments particular elements typically occur in and whether the distribution of elements with respect to each other are complementary or coincidental) can work toward a better or more differentiated definition of these broader categories by working from the bottom up, with the general (and initially vague) goal of fitting individual events and circumstances to these higher level categories.

In the beginning these higher level categories are always provisional. The main point is that one does not impose a hard and fast set of higher order categories on the texts and work down from them. To do so is to impose the analytical and biased categories of the analyst on the data, rather than to let the data speak for themselves. After looking at many different stories in some genre where tentative eidon candidates are defined as eidon varieties, a general structure begins to emerge. For example, different varieties of what would appear to be something we will call an “attack” eidon are found always to form part of an engagement component which, with other eidons appears to involve the most important actions of the story and hence form what can be called the Main Action (MA) component. The eidons of this component can be positioned by a series of rules. In the story above we would have the following rule for the Main Action component:

MA → (Ak (Attack) St (Struggle) Es (Escape) El (Elimination) Dc (Deception)).

This Main Action component, in turn, is part of a higher level category we can call the Engagement Section (E) and is positioned by the rule:

E → (PA (Preliminary Action) MA (Main Action))

This in turn is replaced by a yet higher level segmentation of the story into a motivational section (M) and a Response section (Resp). The Resp consists of an Engagement section and a Resolution section.

Resp → E (Engagement) R (Resolution)

and

Move → M Resp

In sum, eidochronic analysis suggests the existence of a cultural/cognitive system which is available to the story tellers of a particular group. A plot eidon is defined not simply as a narrative event but as part of an

active cognitive system. As such it is a culture-specific, genre-specific natural cognitive unit which can be identified through the distributional study of elements in a large number of folktales collected from the particular culture and genre being analyzed (Colby 1973, Colby and Colby 1981).

The eidon is thus what anthropologists refer to as an emic unit. One of the confusions in discussions about emic and etic units is that an emic unit is often thought to be simply some term or expression uttered by a native, while an etic unit is one created by the researcher. This is to miss the main point of emic analysis, and indeed of phonemic analysis, which was the original linguistic model for the emic-etic distinction. Phonemes are psychologically real sound units which are discovered by the linguist through a distributional analysis of utterances recorded in a phonetic representation using the international phonetic alphabet or some similar system. The fact that English spelling is not a phonemically consistent system of writing attests to the fact that, while phonemes may be psychologically real, they may not be consciously understood as such by native speakers, otherwise written languages would follow a phonemically consistent system of writing which is rare (though Spanish comes close). So also with grammatical units. Grammar is something that is taught in school but rarely noted by a native speaker who is not schooled in grammatical categories. Nevertheless it represents a system of systems that have psychological or cultural reality for individuals and, though argued about by linguists in the finer points, is acknowledged as real and not simply an artifact of the researcher.

So also for eidochronic analysis. While the initial classifications are likely to be wide of the mark, the researcher lumps and splits categories as he works through a large sample of stories until, through successive approximations, he arrives at codings of text which, when defined and analyzed, show distributional regularities that can be described as rules of a narrative grammar.

Grammars can be worked out most easily in highly regular and clearly patterned areas of cultural expression which are limited to particular genres or spheres of expression (or cultural production) in different societies. That is, it is not clear that folktales everywhere can be easily reduced to grammars. Nevertheless, it may be that through the gradual buildup of grammars for those folktale samples that are amenable to study, general principles may emerge that can help in deciphering the more difficult collections of folktales.

A theory of cultural grammars offers the possibility of greater precision and of prediction of a very special kind. It is useful in determining native categorizations of thought and in revealing a certain underlying logic

of behavior and of the way typical events in the world are construed. To be sure, many of the elements that would be found in a cultural grammar would be specific to the particular people who produced the stories analyzed. But there is undoubtedly a more universal logic which ties in with the culture-specific items, and hopefully this universal, pragmatic logic can be brought out gradually through the writing of many culture-specific grammars, particularly of folk narratives.

It is through the universal elements of a narrative grammar, for instance, that the culture-specific elements can be determined in a distributional analysis, because the higher level categories (i.e., motivation and response) tend to be universal, and the analyst requires some higher level reference point to tie the culture specific elements to.

The important thing to remember in the analysis is that the process involves a series of successive approximations and can take many months or years to complete, depending on the complexity and size of the data.

COMPUTER ASSISTANCE FOR EIDOCHRONIC ANALYSIS

Since so much of the analysis involves trial and error categorizations, a computerized system for doing the analysis would be of tremendous value. To date, work on cultural grammars has not utilized computers to the full extent of available programming possibilities. Though we have advanced to the point where the management of large text samples is commonly done with computers, this work does not come close to the state of the art. The existence of sophisticated pattern matching capabilities (i.e., for complex string searches) makes it possible to use, if not a fully automated analytical system, programs that can be a powerful aid in hand analyses. These capabilities, however, have not been very well developed in narrative analysis. For simple tasks there are several DOS-based computer systems available, ranging from *Nota Bene*, an advanced word processor developed for scholarly and scientific writing and research, to *IZE*. The latter operates on key words which the analyst selects (assigned automatically or manually) for each text. These key words are then counted and compared with other key words and on the basis of frequency and distribution *IZE* constructs an outline of the texts. The outline uses the key words, with the most frequent key words making up the higher level categories and the least frequent key words constituting the lower levels. If the initial eidon candidates are characterized by the use of a certain class of words or phrases, they can be automatically keyed in all the texts of the textbase and then examined in an outline which contains other key words that might represent

other eidons. While key words and phrases are rarely sufficient in themselves for the definition of an eidon (or eidon candidate) they are useful in a preliminary study of an eidon's key word environment.

In addition to the organization of texts various other capabilities of IZE facilitate text analysis. Filters, templates, guidelines and ease in exporting and importing texts to other programs are among these useful features. These capabilities are especially helpful in an initial text management phase where text files are examined and classified by genre, or put in some order that might facilitate the analysis.

Once the text management phase has been completed one can apply various programs for content analysis. We are currently developing a program called SAGE (System for the analysis and generation of eidons), initially conceived in a slow prototype system a little over ten years ago, but now being revised and developed for high volume use in Common LISP (available for mainframe computers, Unix workstations and PCs). The program works on any text and is designed to successively approximate hand analysis of content categories. In an appropriate sample of folktales, for example, narratives analyzed by hand are used to begin the process. After the texts are read in, clauses representing eidons which have already been determined are marked. The analyst then begins a series of computer definitions to approximate the hand scoring.

The program has a mapping feature in which the eidons marked by hand (i.e., correctly identified by the analyst and entered into the computer manually) are compared with the eidons that the computer finds and identifies when given the command to apply the eidon rules the analyst has entered into the computer. Thus the analyst can see how well his rules define the eidon concepts he is developing in his mind as he works through the texts and marks those segments he feels instance those concepts. Hits, misses, and false positives are identified and mapped out for him to see. For example, suppose we think that the act of kidnapping the hero or some associate of the hero in a story is a distinct motivational eidon which only appears in the beginning of Eskimo folktales, never in the middle or toward the end; that is, we notice that many Eskimo stories get started by a strong man or ogre kidnapping the hero's wife or child. Then the rest of the story has to do with restoring the member of the hero's family to his or her former situation. Usually this involves a chase or search, a struggle with the kidnapper, and the freeing of the kidnapped victim and a return home. As the analyst reads through a story that contains such a sequence he marks the key clause or clauses of the kidnapping section. He also makes up some rules so that the computer would mark the same section automatically. The problem is that with the rule just defined the computer might mark other sections which do not instance kidnapping. The analyst, in making the rule,

may not have anticipated other text situations that would inappropriately satisfy the requirements of the rule (where a kidnapping did not occur). The analyst thus has to apply the rule and see what other text segments the computer might find that would instance that rule. Also, the analyst must mark other text segments in other stories where he knows that kidnapping occurs to see how well his rule, or set of rules is operating.

After a series of successive approximations the procedure should come sufficiently close to matching the hand analysis so that new (un-analyzed) narratives can be read in for this kind of semiautomatic analysis.

The program can retrieve sentences or clauses by number or by rule application. It can retrieve them by eidons (i.e., sentences or clauses marked as eidon members), and keep score of how well the computer rule system is matching the hand analysis of the analyst. One can define rules which involve sequences of lexemes (i.e., a word string) or sequences of semantic features. This last ability is a key part of the system and is a major point of differentiation between SAGE and text analysis systems currently available commercially for the PC. Semantic features for the word 'table' could be of several different types including classificatory terms: object, furniture, artifact; associated objects and activities: chair, eat, meal, etc., or grammatical usage (noun, verb, etc.).

A better understanding of how SAGE works might best be attained through a summary of its commands and procedures.

The major functions are accomplished with the "top level" commands:

- | | |
|--------------------|--|
| (init filename) | INITialize — reads in a source of TEXT (one story/paragraph) for analysis. |
| (rs) | Retrieve-Sentences — will print out sentences which adhere to various conditions. This routine prompts the user for some input. |
| (re) | Retrieve-Eidons — will print out sentences which have been flagged with eidons requested by the user. This routine prompts the user for some input. |
| (me) | Mark-Eidon — allows the user to flag sentences with a given eidon. This routine prompts the user for some input. |
| (maps *tag *eidon) | This will create a table of hits, misses, clashes and false positives for the current set of sentences. Tag and eidon are optional. If they are not supplied, the user will be queried for them. (If you wish to specify an eidon directly in the command, you MUST specify the eidon first.) |

You will also be queried for other information (e.g., file or terminal output, etc.).

Existing files are loaded with (tagsin *file). This command will load in a set of tags, features, rules, and morphological variants from the specified file. It assumes a default extension of "tags" for that file. The user is not required to specify a file; in this case, the file name is assumed to be the same as given in the last INIT command. (This command actually just loads in the contents of the specified file, whatever they may be.)

The user can define features, rules, tags, morphs, or lexemes with the command DEFINE. For example, a rule would be defined like this: (define ^instrumental to use# power). The prefix/suffix characters specify what the following symbol will be. In the following example the \$ signifies a tag, which is the computer approximation to an eidon: (define \$power ^instrumental power prestige). No prefix for the word immediately following "define" indicates that the word is being defined in terms of its features (which follow the word being defined).

You may "view" the definitions of tags, features, rules, and morphological variants by using the VIEW command. E.g., (view \$power) will display the definition of the power tag. Similarly for the others.

Rule patterns have the following components:

symbol	:: this is to be matched literally.
symbol#	:: any of the "morphs" of symbol is to be matched literally.
@	:: must match one of the lexemes associated with the feature symbol.
(@X @Y)	:: one lexeme must possess all of the features given.
X Y	:: is the "tie," i.e., X is followed immediately by Y. Note that the default is for any number of lexemes to be "passed by" when matching the pattern of the rule. E.g., the pattern "TO PLAN# POWER" means that TO must be followed immediately by some morph of PLAN which is to be followed (possibly skipping some of the lexemes in the clause) by some lexeme with a POWER feature.
[X ! Y ! ... !Z]*	:: order independence, i.e., X and Y but in either order. Processing is left to right. This implies that if two subsentences potentially match the same lexeme in the clause, the 1st subpattern in the rule will match, excluding a match

by the second. This may affect the rule's behavior in the following way. Suppose the clause is: FOB BAR JAK QOD and the rule pattern under consideration is:

[@A ! @] with these feature definitions:
 (define @A BAR JAK)
 (define @B BAR)

The above rule will not match the clause because of the left to right processing. The @A feature will match the BAR lexeme. BAR is then "removed" from the clause—it will not be considered for further matches. Since there is nothing left for the @B feature to match, the rule pattern fails to match the clause. If, however, @B were defined as:

(define @B BAR FOB)

then the rule pattern would match the clause. :: exclusive or (disjunction). As with order independence "[X ! ... ! Z]" processing is left to right. The first given subpattern to match will cause this entire pattern to match.

{X ! Y !...! Z}*

*X, Y, etc. can be any legal pattern.

"!" separates the patterns.

Nesting of operations is okay but discouraged because it may be computationally expensive.

Editing and saving are the remaining important functions of the program. The user may edit online definitions of tags, rules, features, and morphological variants by using the EDIT command. One can, for instance, erase (remove) an item from memory, add new-members to an item, delete members from an item, or replace old (member) with new (member) for the item. (Note that this edits the online definitions only, NOT the definitions on files.)

Tags, rules, etc. may be saved with (tagout *file). This will write out all of the tags, features, rules, and morphological variants into a file suitable for use by TAGSIN. The filename is the same as for TAGSIN. If a file exists with the same name, that "old" file will be overwritten.

CONCLUSION

With SAGE we hope that it will be possible to analyze large numbers of folktale texts or other kinds of formulaic material, preferably of an oral

tradition, to determine plot units that have special cultural reality for the particular people who have produced the texts. To reiterate the condition for analyzing plot grammars, it is necessary that the sample of texts be geographically bound to a particular language using group of people and that it consist of the same genre and general time period. With these restrictions, and if the sample is sufficiently large (numbering at least over fifty and preferably twice as much) it should be possible to eventually work out a plot grammar. This grammar can go a long way towards bringing out the "cultural logic" of a culture using group of people. Once a sufficiently varied and large number of such grammars have been worked out, always working from the bottom up rather than the top down, we can move to the next major challenge in cultural analyses, how to decode and represent the cultural logic that people use to interpret events and behavior in their world.

ACKNOWLEDGMENTS

This paper draws upon sections of a paper read at the meetings of the American Anthropological Association, Washington D.C., November 1989: *Computer Studies of Stress and Health Discourse* by B. N. Colby, S. Mishra, L. Milanese, and D. P. Adams.

REFERENCES

- Bateson, G.
1936 *Naven*. Stanford University Press.
- Colby, B. N.
1973 "A Partial Grammar of Eskimo Folktales" *American Anthropologist* 75 (3) 645-662.
- Colby, B. N. and L. M. Colby
1981 *The Daykeeper; the Life and Thought of an Ixil Diviner*. Cambridge, MA: Harvard University Press.
- Faris, J. C.
1972 *Nuba Personal Art*. London: Duckworth.
- Mehan, H.
1979 *Learning Lessons: Social Organization in the Classroom*. Cambridge, MA.: Harvard University Press.
- Propp, V.
1968 *Morphology of the Folktale*. 2nd rev. edition, L. A. Wagner (ed.) Austin: University of Texas Press.
- Roberts, J. M.
1964 *The Self-Management of Cultures*. In *Explorations in Cultural Anthropology; Essays in Honor of George Peter Murdock*. Ward Goodenough (ed.). New York: McGraw-Hill.
- Romney, A. K., S. C. Weller, and W. H. Batchelder
1986 "Culture as Consensus: A Theory of Culture and Informant Accuracy." *American Anthropologist* 88 (2) 313-338.

Spencer, R. F.

1959 *The North Alaskan Eskimo*. Washington D.C. Bureau of American Ethnology.

Watt, W. C.

1986 "Canons of Alphabetic Change" *Studies in the Cognitive Sciences* 38. Irvine, CA.: School of Social Science, University of California, Irvine.