**Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis**

Kathleen Carley

*Sociological Methodology*, Vol. 23 (1993), 75-126.

Stable URL:
http://links.jstor.org/sici?sici=0081-1750%281993%2923%3C75%3ACCFTAA%3E2.0.CO%3B2-1

*Sociological Methodology* is currently published by American Sociological Association.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/asa.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

For more information on JSTOR contact jstor-info@umich.edu.

©2003 JSTOR

http://www.jstor.org/
Fri Nov 21 12:56:25 2003

3

# CODING CHOICES FOR TEXTUAL ANALYSIS: A COMPARISON OF CONTENT ANALYSIS AND MAP ANALYSIS

*Kathleen Carley**

*Content and map analysis, procedures for coding and understanding texts, are described and contrasted. Where content analysis focuses on the extraction of concepts from texts, map analysis focuses on the extraction of both concepts and the relationships among them. Map analysis thus subsumes content analysis. Coding choices that must be made prior to employing content-analytic procedures are enumerated, as are additional coding choices necessary for employing map-analytic procedures. The discussion focuses on general issues that transcend specific software procedures for coding texts from either a content-analytic or map-analytic perspective.*

## INTRODUCTION

Are female role models for children changing? Are concepts like freedom and love universal invariants across humans or are there cultural differences in what is meant when such concepts are used? Does the language of a group change as it becomes a profession? When do the advertisements of companies or the speeches of politicians signal real shifts in intent or focus and when do they reflect just shifts in the rhetoric? Questions such as these can be addressed by analyzing texts. Indeed, much social information is in the form of

*Carnegie Mellon University

texts such as answers to interview questions, books, articles, newspaper clippings, and essays. Such texts are a rich and valuable resource for understanding social behavior, but they are very difficult to analyze. Still, many questions are best addressed by examining texts, and in many cases an extremely large number of texts.

Textual analysis is vital to social science research, and a wide range of techniques have emerged. These include computational hermeneutics (Mallery 1985; Mallery and Duffy 1986), concordance analysis (for example, see Young, Stevenson, Nicol, and Gilmore's concordance of the Bible (1936) or Ellis's concordance of the works of Shelley [1968]); content analysis (see Weber 1984 for an overview; for a slightly different approach, see also Neuman 1989); conversational analysis (Sacks 1972); data-base techniques for conducting ethnographic and qualitative studies (vol. 7 of *Qualitative Sociology* 1984 was devoted to this issue; see also Blank, McCartney, and Brent 1989; Sproull and Sproull 1982; Qualis 1989); discourse analysis (Stubbs 1983; Polanyi 1985; Sozer 1985); linquistic content analysis (Lindkvist 1981; Hutchins 1982; Roberts 1989); semantic grammars (Franzosi 1990*a*, 1990*b*); protocol analysis (Ericsson and Simon 1984); procedural task analysis (VanLehn and Garlick 1987); proximity analysis (Danowski 1980, 1982, 1988; Van Meter and Mounier 1989); and story processing (for example, see Lehnert and Vine 1987; Mandler 1984; Dyer 1983). Each of these techniques has advantages that enable the researcher to consider the text in a certain fashion. While it is not the purpose of this paper to describe and contrast this plethora of approaches in detail, a few covering comments are appropriate.

Textual analysis techniques tend to be extremely time-consuming to apply. Many of the techniques are not automated. As a consequence, many researchers resort to handling only a few texts in a highly qualitative fashion. Despite their richness, such analyses typically lack precision and inferential strength. Alternatively, researchers engaged in more empirical investigations often "reinvent the wheel"—for example, developing yet another word counting program. Of the automated and semiautomated techniques, most are ad hoc, developed for a specific research group or applied to a specific problem, and they are not generalizable to different groups with different problems. Still other automated techniques, particu-

larly those that focus on generating concordances or retaining the original text, often have substantial associated processing and storage costs. Many of the techniques developed have been applied only to a single text. Consequently, issues centering around the comparison of multiple texts have not been addressed. Finally, even those techniques that are automated and robust have rarely been designed to be used in conjunction with other methodological tools such as statistical packages.

In response to this morass of problems, attention has often been directed to the characteristics of a particular analysis technology rather than to the underlying choices that a researcher must make in order to use that technology. Such choices, however, are critical. They can dictate what technology is appropriate and affect the types of findings possible.

Within the social sciences, the dominant solution to textual analysis problems has historically been content analysis (Fan 1988; Namenwirth and Weber 1987; Garson 1985; Stone et al. 1968). Content analysis enables quantitative analysis of large numbers of texts in terms of what words or concepts are actually used or implied in the texts. Although extensively used, this approach has met with only limited success for a variety of reasons, including lack of simple routines, time-consuming data preparation, difficulties in relating textual data to other data, and a lack of a strong theoretical basis.[1] By taking a content-analytic approach, the researcher has chosen to focus on largely isolated concepts. This choice vastly simplifies coding, and it allows the researcher to address a variety of questions: How does the distribution of word usage change over time? Is the style of two texts similar in terms of the proportions of types of words? However, as will be demonstrated, the focus on concepts implicit to traditional content analysis often results in an overestimation of the similarity of texts because meaning is neglected.

An important class of methods that allows the researcher to address textual meaning is map analysis. Where content analysis typically focuses exclusively on concepts, map analysis focuses on concepts and the relationships between them and hence on the web

[1]For a more extensive discussion of the difficulties and advantages of this approach, see Weber 1984; Roberts 1989; Namenwirth and Weber 1987; and Neuman 1989.

of meaning contained within the text.[2] While no term has yet to emerge as canonical, within this paper the term *map analysis* will be used to refer to a broad class of procedures in which the focus is on networks consisting of connected concepts rather than counts of concepts. Alternate terms that have been suggested include cognitive mapping, mental model analysis, frame analysis, cognitive network analysis, scheme analysis, relational analysis, meaning analysis, and relational meaning analysis. Map analysis is a family of techniques, all of them emphasizing situated concepts and the relationships between them but varying in whether they look at relationships from a simple semantic (Carley 1986; forthcoming), proximal (Danowski 1982), or linguistic (Roberts 1989; Dyer 1983) perspective.[3] Like content analysis, map analysis enables quantitative analysis of large numbers of texts.[4] But unlike content analysis, map analysis has a strong theoretical base stemming from a variety of research: the construction of meaning (Fauconnier 1985); mental models (Johnson-Laird 1983; Gentner and Stevens 1983); knowledge representation (for overviews see Bobrow and Collins 1976 and Brachmand and Levesque, 1985); conceptual structures (Sowa

[2]Several researchers have employed techniques based on content analysis to extract "maps" limited in form to a particular type of relationship. For example, particular models of communication, sentence structure, or story format are predefined by the researcher and then texts are analyzed by coding the content into these predefined categories with predefined relations. The results of these analyses are typically treated in terms of "counts" of the predefined patterns that occur, not in terms of the web of meaning within the text. The map procedure described herein enables the researcher to extract these specialized "maps," but also to extract more generalized webs. A second example is "contingency analysis," where the researcher looks for the co-occurence of two concepts in some unit of text (such as 120 words) (Holsti 1954; Osgood 1959). This technique is similar to Danowski's (1980, 1982) proximity analysis. Unlike map analysis, these techniques focus on counts (in this case, of paired concepts) and not on the overall pattern of concepts.

[3]Linguistic approaches are not typically viewed from this perspective because of their emphasis on story lines, the temporality of events, and/or procedures. Nevertheless, texts coded from a linguistics perspective can be represented and analyzed as maps, although this is rarely done. Rather, the focus is often on translation, abstraction, and story regeneration. This difference in emphasis, as well as the additional syntactic difficulties in coding texts from this perspective, often lead researchers to consider this approach as fundamentally different from the simpler semantic and proximity approaches.

[4]Linguistic approaches, unlike either simpler semantic or proximity approaches, have rarely been used on large numbers of texts.

1984); schemas (Mandler 1984); social language usage (Dietrich and Graumann 1989; Kaufer and Carley 1993); personal knowledge (Polanyi 1962); social knowledge (Cicourel 1974); and cultural truth (Romney, Weller, and Batchelder 1986). Carley and Palmquist (1992) or Carley (1988) provide additional details on the theoretical base for the map analysis and procedures followed to code data that employ a simple semantic approach using one possible software package.

By taking a map-analytic approach, the researcher has chosen to focus on situated concepts. This choice increases the complexity of the coding and analysis process, and places the researcher in the position where a number of additional choices must be made regarding how to code the relationship between concepts. As I will demonstrate, these additional choices can also affect the degree of similarity found between texts. Because they focus on situated concepts, map-analytic techniques produce networks of concepts that can be examined at both a graphical and a statistical level. Given a set of maps and using either graphical or statistical techniques, we can ask a variety of questions: Do individuals' mental models become more complex over time (i.e., do they come to contain more concepts, more relationships, or a higher ratio of relationships to concepts)? Do concepts denoting emotions have different structural properties than do other concepts (i.e., do they have more relationships, different types of relationships)? Because map analysis makes it possible to examine the data graphically and statistically, the researcher can stay close to the text and so augment qualitative techniques and capture the precision and inferential ability of quantitative techniques.

Choices, such as whether to focus on isolated or situated concepts, with all the attendant options, transcend specific software procedures for coding texts and, in general, should be made prior to choosing a software package so as to ensure that the chosen software does not constrain the choices for the researcher. This paper discusses the coding choices that the researcher must make when taking either a content-analytic or map-analytic approach. This will lead to a discussion of the relative advantages of these highly useful techniques. In addition, the discussion will briefly touch on how to compare texts given the way in which they were coded using either content-analytic or map-analytic techniques. Comparison of coded texts, however, is a major issue and, for the most part, is well beyond

the scope of this paper. Rather, this paper centers on those choices and issues that the researcher should be aware of prior to engaging in the coding and analysis of texts.

Clearly, all of the other approaches for textual analysis previously alluded to require the researcher to make choices that affect how the texts are interpreted and the potential results. As a consequence, each approach has its advantages and disadvantages. In this paper, however, the primary emphasis will be on content analysis and map analysis as they both share certain characteristics: They require the researcher to address many of the same coding and theoretical questions; they are automated or semiautomated; they facilitate automated comparisons of large numbers of texts; and they have a wide range of usage and applicability across the social sciences and humanities. Moreover, in terms of coding choices, the contrast of content and map analysis is particularly interesting as map analysis subsumes content analysis. I will explicitly discuss the additional analytical power of map analysis over content analysis and the associated additional analysis requirements.

Discussions of the coding and analysis of texts often include information on the procedures and tools available. In this paper, all coding and analyses are done using various procedures from the MECA toolkit (Carley 1990).[5] By analogy with statistical packages, MECA is a general "package" or "toolkit" for analyzing texts that contain a set of techniques—some of which retain the richness of more qualitative approaches, automate or semiautomate coding procedures, and enable quantitative analysis; many of which generate data in a form that can be used with standard statistical packages; and all of which can be used in conjunction with a qualitative or quantitative approach. Nevertheless, this paper is not about MECA, nor is its purpose to describe the tools.[6] Rather, it centers on questions that transcend particular tools: What is the difference between content and map analysis? What are the relative advantages of the two approaches? What are the issues underlying their usage? Be-

[5]The MECA tools are written in C and are available for many PC, Macintosh, and UNIX platforms. Additional information on MECA is available from the author. Many tools other than MECA are available for content analysis and map analysis.

[6]Additional details on MECA are available in Carley 1990*a*.

cause of this emphasis, details on what procedures were used are, when necessary, relegated to footnotes and the appendix. As noted, the ensuing discussion presents a series of coding choices that the researcher must make when analyzing texts. The MECA tools do not make these choices but allow the researcher to make any set of choices for either content or map analysis.

## 1. CONTENT ANALYSIS

Content analysis focuses on the frequency with which words or concepts occur in texts or across texts. This approach has been used to examine a variety of topics such as conceptual shifts in presidential addresses (Sullivan 1973) and cultural changes (Namenwirth and Weber 1987). The basic idea is to take a list of concepts and a set of texts and then simply count the number of times each concept occurs in each text. Differences in the distribution of counts across texts provide insight into the similarities and differences in the content of the texts.

Before continuing, it will be useful to define the term *concept*. A concept is a single idea, or ideational kernel (Carley 1986), regardless of whether it is represented by a single word or a phrase. Examples of concepts are "friends," "textual analysis," and "likes to play golf."

### 1.1. *Coding Choices*

There are a large number of choices that must be made by the researcher when doing content analysis—e.g., level of analysis and generalization. These choices—those discussed below as well as many others—affect what results are achieved, the interpretation of the results, and whether it is easy to automate the coding process. Further, it is important that the researcher be aware that not only are these choices being made because of potential methodological bias but also because such choices require semantic and cultural interpretations of the data (Cicourel and Carley 1990). Moreover, the researcher should be aware when choosing an automated or semi-automated procedure for conducting content analysis that many of these choices already may have been made.

1.1.1. *Level of Analysis—What Constitutes a Concept*
Different results will emerge if single words, as opposed to phrases, are used in the coding. Using single words is particularly useful if the researcher wants to contrast the results in a specific text or type of text with general usage. In contrast, phrases are useful when the researcher is interested in capturing broad-based concepts or terms of art in a particular sociolinguistic community (e.g., the phrase "Standard Operating Procedure" or "SOP" in the military).

1.1.2. *Irrelevant Information*
Should irrelevant information be deleted, skipped over, or used to dynamically reassess and alter the coding scheme (thus effectively treating no information as irrelevant)? Some methods for dealing with irrelevant information—such as deleting it from the text—can facilitate a more automated approach to content analysis by (1) minimizing storage and processing costs; (2) facilitating automatic coding of texts, and (3) generating a simplified text that can be visually inspected. Procedures for deleting concepts are particularly useful when the researcher wants to eliminate special notation, proper names, articles, and so forth. When there are large sections of text, such as paragraphs, that are simply irrelevant, it is more efficient to simply delete them with a text editor. There are two main difficulties associated with deleting concepts: embedding and interpretation. If the concept to be deleted is embedded in another concept (e.g., the concept "in" is embedded in "fits in with the hall"), then deletion of the embedded concept may interfere with locating those concepts that are to be coded. Thus deleting all occurrences of "of" may make it difficult to locate the concept "fortress of solitude." Deleting unwanted concepts also may make the resultant text difficult to read. Consider the following text:

> Killishandra and Ezra tailed the merchant through many corridors in order to find the robbers' lair. Unfortunately, he managed to elude them.

Now, let us delete all proper names, pronouns, conjunctions, articles, prepositions, and notation.[7] The resultant text is:

[7]This can be done, for example, by applying the MECA program DELCON to a file containing the text.

tailed merchant through many corridors find
robbers lair Unfortunately managed elude

Thus in many cases researchers may want to keep copies of both the processed and unprocessed text.

Determining what information is irrelevant is in itself a choice that must be made by the researcher. There are no set standards for defining information as irrelevant. Common standards are deletion of text not central to the research question and deletion of articles. A focus on "action" may lead to the deletion of proper names, pronouns, etc., as in the foregoing example.

1.1.3. *Predefined or Interactive Concept Choice*

When coding a text the researcher must choose between using a predefined set of concepts or developing a list of concepts incrementally during the process of coding. Unless the researcher is interested in coding all of the concepts that occur in the text, having a predefined list is a prerequisite for automated coding. Having a predefined list of acceptable concepts can be done by specifying either which concepts are to be coded or which are to be deleted.

1.1.4. *Level of Generalization*

Are all explicit concepts to be coded exactly as they occur in the text or are they to be recoded in some altered or collapsed form? Coding concepts as they occur in the text facilitates automation but at the cost of cross text comparability. Generalization—for example, when tense is ignored or all pronouns are converted to proper nouns—admits greater comparability across texts but at the cost of lost information. Choosing the right level of generalization is in many ways an art form dictated both by theoretical concerns and by the type of analysis in which the researcher wishes to engage. Some of the concerns in deciding the appropriate level of generalization can be illustrated using the following two texts:

*Text A:* Killishandra and Ezra tailed the merchant through many corridors in order to find the robbers' lair. Unfortunately, he managed to elude them.

> *Text B:* Blair and Tony followed the store-owner across town to find the thiefs. Unfortunately, he managed to lose them.

Both passages might represent to the researcher the idea of "failed search." A content analysis using only explicit concepts and no generalization reveals that the only concepts the two passages have in common are "and," "the," "to," "find," "the," "unfortunately," "he," "managed," and "them." None of these shared concepts reveal the idea of "failed search." Now, let us allow generalization. The first thing to note is that there exists no single word or phrase in these passages that corresponds to the idea "failed search." Rather, this is a gestalt impression achieved by reading both sentences in each passage. Thus direct translation of words or phrases into "failed search" will not reveal any similarity in the texts. The general principle is that the more complex the concept that one is trying to generalize to the less likely it is that specific synonyms will appear in the text. Thus to extract more complex concepts, the appropriate approaches are those that involve locating the co-occurrence of simple concepts or those that use rules to translate co-occurring concepts into new concepts. (An example of this latter approach can be seen in Fan 1988.)

A more bottom-up approach that separately locates the concept "failed" and "search" may be more successful. Make the following translations: the concepts "tailed," "find," "followed" into the concept "search"; the concepts "elude" and "lose" into "failed"; the concepts "merchant" and "store-owner" into "businessman"; and the concepts "robbers" and "thiefs" into "crooks." The resultant texts would appear as:[8]

> *Modified Text A:* Killishandra and Ezra search the businessman through many corridors in order to search the crooks' lair. Unfortunately, he managed to failed them.

> *Modified Text B:* Blair and Tony search the businessman across town to search the crooks. Unfortunately, he managed to failed them.

[8]This translation was accomplished by employing the MECA program TRANSLATE. See appendix for additional information.

A content analysis of these modified texts reveals that these passages share not only the concepts described above but also the concepts "search," "businessman," "crooks," and "failed." The main problem is overgeneralization—that is, using categories of concepts that are so vague that important semantic distinctions are lost. For the above example, this might occur if all proper-person-names, pronouns, and types of people were converted to the concept "person," and all location names were translated to the concept "place." Applying this translation to the modified Text A results in:

> *Remodified Text A:* Person and person search the person through many place in order to search the persons' place. Unfortunately, person managed to failed person.

Such overgeneralization not only produces ludicrous sounding texts but can also make texts that are different appear identical.

1.1.5. *Creation of Translation Rules*

In order to systematically generalize concepts in the texts, it is necessary to create a series of "rules" or a thesaurus that translates less general concepts into more general ones. How these rules are created depends on the researcher's goals. There are two generic approaches: using an actual thesaurus and using a specially constructed thesaurus. Using an actual thesaurus has the advantage that the information is predefined. Such thesauruses are now available on line and their data bases can be easily converted to the form required for textual analysis. The disadvantage of this approach is that special slang or meanings within a particular sociolinguistic environment are likely to be missed. An alternative is to construct a special purpose thesaurus for the texts that will be analyzed based on detailed analysis of a sample of the texts. This guarantees that the thesaurus contains information to deal with the peculiarities of the particular sociolinguistic environment. Since only a sample of the texts is used, however, some possible synonyms may be missed.

1.1.6. *Level of Implication for Concepts*

Are the texts to be coded in terms of which concepts are explicitly present or in terms of which concepts are implied? Locating implicit

knowledge goes beyond generalization as it involves not just mapping one concept onto another, as in the case of generalization, but determining from a set of concepts what other concepts are missing. The ability to extract implicit concepts is vital to much research as meaning is lost when only explicit concepts are used (Ogilvie, Stone, and Kelly 1982; Woodrum 1984), or as Merleau-Ponty (1964, p. 29) contends: "The totality of meaning is never fully rendered: There is an immense mass of implications, even in the most explicit of languages." The level of implication affects the type of results, their interpretation, and the possible level of automation. For example, if only explicit concepts are used, then texts can be compared in terms of differences in style (i.e., word usage), and a completely automated procedure can be followed. In contrast, the use of implied concepts may allow the researcher to compare texts in terms of underlying shared meanings and social knowledge, but it may make it more difficult to completely automate the process. Locating implicit knowledge in the text is quite difficult, and limited success has been achieved using knowledge of the sociocultural environment that generated the text (Carley 1988) and the requirements of the story line (Schank and Riesbeck 1981).

1.1.7. *Existence or Frequency*

Should texts be compared in terms of simply whether or not a concept occurs or in terms of how frequently the concept occurs? While simple occurrence-based comparisons simplify discussions of co-occurrence and eliminate frequency biases due to syntactic requirements, frequency-based comparisons make possible discussions of saliency and emphasis. This decision can be put off to the analysis stage. That is, the researcher can code the data in terms of frequencies and then later collapse to existence.

1.1.8. *Number of Concepts*

How many concepts should be used in the analysis? This choice is not independent of the previous set of choices. In general, a total of approximately 100 to 500 concepts seems sufficient to code the knowledge on a specific topic in any sociolinguistic environment. For example, Carley (1984) used 217 concepts for describing talk about tutors among students engaged in a decision-making context; Palmquist (1990) used 212 concepts for describing talk in the conventional class-

room and 244 concepts in the nonconventional classroom; Cicourel and Carley (1990) used 174 concepts for describing the content of a story and its recall by children; Carley and Kaufer (forthcoming) used 310 concepts for describing information on drama and comedy; and Thomas (1991) used 395 concepts for describing the portrayal of robots in science fiction books. A set of 100 to 500 concepts generally represents a slight level of generalization, which means a decrease from the absolute number of concepts in the texts. Concordances, for example, which can contain every word in the text, may list thousands and even tens of thousands of words. Coding texts with many fewer concepts, such as less than 25, tends to obfuscate meaning. Coding texts with many more concepts, such as more than 1000, tends to prevent text comparison (unless one is interested in separating exact word usage). That is, while a total of 100 to 500 concepts seems sufficient to capture many of the nuances and individual differences within texts, it is still a small enough number that some generalization and some comparison in terms of shared meanings is possible.

These eight coding choices serve to illustrate the types of questions that the researcher must address when coding texts using a content-analytic technique. The choices that are made affect what tools are needed by the researcher and even the likelihood that there will be a tool available.

### 1.2. *Advantages and Limitations*

Content analysis is a fairly versatile technique that can be applied to a wide range of texts. It lends itself to automation, and, as I mentioned above, it is easy to create automated tools for locating those concepts that are explicit in the text. As a result, there are a large number of programs for doing this, including basic UNIX utilities like *grep,* large general systems such as the General Inquirer (Stone and Cambridge Computer, 1968; Stone et al. 1968), and assorted smaller or special purpose programs developed by individual researchers, such as those by Garson (1985) or Carley (1990). Such tools differ from each other primarily in the length of text that they can analyze, the complexity of the concepts they can locate, the number of texts they can process at once, and the hardware on which they run.

When the research question requires the extraction of informa-

tion that is implicit only in the text, content analysis fares less well. Furthermore, it is doubtful that a simple solution within the scope of content analysis will be found. Indeed, the search procedures to explicate implicit information has been one of the central problems faced by researchers in artificial intelligence who are interested in locating the deep structure or complete understanding of texts (and in particular stories) (Schank and Riesbeck 1981; Bruce and Newman 1978; Lehnert 1981; Dyer 1983). When implicit as well as explicit concepts are coded, it is possible to extract from the text a richer definition of meaning but at the cost of ease of automation (at least with current technology). One reason for this is that the extraction of implicit concepts often requires impressionistic judgments (Holsti 1969; Krippendorf 1980), which when made by humans frequently result in coding mistakes such as errors of omission (see Carley 1988).[9]

Techniques developed by researchers in artificial intelligence for locating implicit meaning—such as conceptual dependency-based scripts (Schank and Abelson 1977), plot units (Lehnert 1981), and techniques for coding affect (Dyer 1983)—provide wonderfully detailed representations of the text's implicit content. These techniques, however, require such vast amounts of preanalysis of the texts by the researchers that they have been applied to only a few sample texts (i.e., to two or three) and thus, though they illustrate the power of the approach, are not practical tools for the researcher interested in coding and comparing vast numbers of texts. The need to analyze vast numbers of texts is the problem more commonly faced by social scientists. Thus, within the social sciences, it has become more common to use dictionaries to locate implicit knowledge. The compilation of such dictionaries or thesauruses is fraught with difficulties. (For additional details see Namenwirth and Weber 1987, ch. 8, and Carley, forthcoming *a*.) Even so, they tend to increase reliability (Grey, Kaplan, and Lasswell 1965; Saris-Gallhofer, Saris, and Morton 1978). One such difficulty is trying to distinguish between homographs—words that have the same written form but a different origin and meaning, as in record (transcribe), record (vinyl disk), and record (best past achievement). Rule-based approaches

[9]For a more detailed discussion of errors and the procedures for preventing or detecting them via a grammatical approach, see Franzosi 1990*c*.

often are used (Dunphy, Bullard, and Crossing 1974; Namenwirth and Weber 1987; Fan 1988; Carley 1988) to overcome this and other difficulties. Another difficulty is that construction of dictionaries, thesauruses, and rule sets is a time-consuming process.

Even when implicit concepts are used, the main strength of content analysis is that it can be (although it is not always) totally automated and so applied to vast numbers of texts, thus making possible empirical cross-subject and cross-time comparisons. There is, however, a fundamental theoretical problem with simply extracting concepts. That is, the presence of concepts may not be sufficient to denote meaning as people can use the exact same words, even when they are not homographs, and yet mean very different things. In part, this is because the same words used in different contexts have very different implications. This difficulty can be addressed using procedures such as contextual content analysis (McTavish and Pirro 1990). However, even given functionally identical contexts, the problem still may arise. Consider the following texts, representing the views of two students about scientists:

> *Student A:* I found that scientists engage in research in order to make discoveries and generate new ideas. Such research by scientists is hard work and often involves collaboration with other scientists which leads to discoveries which make the scientists famous. Such collaboration may be informal, such as when they share new ideas over lunch, or formal, such as when they are co-authors of a paper.
>
> *Student B:* It was hard work to research famous scientists engaged in collaboration and I made many informal discoveries. My research showed that scientists engaged in collaboration with other scientists are co-authors of at least one paper containing their new ideas. Some scientists make formal discoveries and have new ideas.

Clearly the meanings of the texts are different. Yet this difference may not be revealed by content analysis. These two texts were

TABLE 1
Content Analysis Comparison of Two Texts

| Concept | Student A | Student B |
|---|---|---|
| I | 1 | 1 |
| scientists | 4 | 4 |
| research | 2 | 2 |
| hard work | 1 | 1 |
| collaboration | 2 | 2 |
| discoveries | 2 | 2 |
| new ideas | 2 | 2 |
| formal | 1 | 1 |
| informal | 1 | 1 |
| coauthors | 1 | 1 |
| paper | 1 | 1 |

coded using only the explicit concepts: I, scientists, research, hard work, collaboration, discoveries, new ideas, formal, informal, coauthors, and paper.[10] Table 1 contains the result. As this table illustrates, each concept is used exactly the same number of times. Not only is there no difference in which concepts are used, there is also no difference in emphasis (frequency with which the concepts are used). This is not to imply that content analysis can never locate differences in meaning. In fact, content analysts typically try to locate meaning shifts over time or across people by analyzing differences in the frequency with which different concepts are used across texts using factor and/or cluster analysis (Namenwirth and Weber 1987; Gallhofer and Saris 1988), analysis of variance (Potter 1982), or other statistical techniques. As noted by Namenwirth and Weber (1987, p. 195), content-analytic procedures are appropriate and work well at macro levels where the interest is in broad cultural concepts but seem less effective at the micro level where psychological or cognitive phenomena need to be examined. The comparison of the texts by Students A and B is at this micro level, and as Table 1 illustrates content analysis reveals no difference.

Why is the difference in meaning not revealed by this analysis? One reason is that individuals may differ in their definition of

[10]The MECA program CMATRIX2 for doing content analysis was used. See the appendix for additional information.

words. For example, Student A seems to have a more elaborate definition of "collaboration," involving both formal and informal interactions; Student B seems to define "collaboration" exclusively in terms of coauthoring. Or to draw on another rhetorical context, one political analyst's definition of a military crisis might require the presence of nuclear weapons, while another's might require only projected differences in available forces. A second reason is that the same word may be employed in different syntactic locations. For example, Student B uses the concept "scientists" in the first sentence as an object-noun in the prepositional phrase "to research famous scientists," whereas later, in the third sentence, Student B uses the concept "scientists" as the subject of the sentence "scientists make formal discoveries." A third reason is that the same word may be used in multiple semantic contexts. For example, these students are discussing scientists relative to the context of a class assignment and are focusing on generalities. Within the context of an NSF panel review, the discussion of scientists would assuredly be quite different.

Under each of these conditions, the difference in meaning is revealed not by differences in which concepts are used but by differences in the relationships between concepts. In terms of definitions, Student A links "collaboration" to "formal" and "informal" and links "formal" to "coauthors" and "paper," whereas Student B directly links "collaboration" to "coauthors" and "paper." In terms of syntax, Student A has a link from "research" to "discoveries" to "scientists," whereas Student B has a link from "research" to "scientists" to "discoveries." In terms of semantic context, Student A links "scientist" to "research" and Student B links "I" to "research." Thus at the micro level, particularly when the researcher is interested in differences in meaning, it is necessary to examine how people actually interrelate concepts, rather than just how concepts covary across texts, in order to preserve the semantic and syntactic structure of the text (Carley 1986; Carley and Palmquist 1992; Roberts 1989).

## 2. MAP ANALYSIS

Map analysis compares texts in terms of both concepts and the relationships between them. This approach has been used to examine a variety of topics, including voting behavior (Carley 1986), shifts in

views toward research writing (Palmquist 1990), children's memory of stories (Cicourel and Carley 1990), across country and time differences in the World Bank's strategy (Saburi-Haghighi 1991), and shifts in the cultural perception of robots as evinced by science fiction writers (Thomas 1991). The basic idea is to take a list of concepts and a set of texts and then to determine for each text whether these concepts occur in the text as well as the interrelationships between those concepts that do occur. Frequency information also may be gathered. Differences in the distribution of concepts and the relationships among them across texts provide insight into the similarities and differences in the content and structure of the texts.

Before continuing, the terms *concept type, relationship,* and *statement* need to be defined. Concepts can be organized into categories or types such that one specific concept is an instance of another more general type; for example, "Corwin" and "Cassandra" are instances of people. A relationship is a connection between concepts, whereas a statement is two concepts and the relationship between them. Different relationships have different *strength, sign, direction,* and *meaning*. These characteristics are defined in detail in Carley and Palmquist (1992) and Carley (1988). It will suffice here to simply provide illustrations of each of these characteristics. Consider the following statements:

1. Cassandra loves Corwin.
2. Corwin loves Cassandra.
3. Peter likes Cassandra.
4. Corwin does not like Peter.
5. A table has 4 legs.

Statements 2 and 3 differ in the strength of the relationship—"loves" being stronger than "likes." Statements 3 and 4 differ in the sign of the relationship—"likes" indicates a positive relationship whereas "does not like" indicates a negative relationship. Statements 1 and 2 differ in the directionality of the relationship—in statement 1 the relation goes from Cassandra to Corwin and in statement 2 the relation goes from Corwin to Cassandra. Finally, statement 5 contains a different type of relationship than do all other statements—the relationship in statement 5 is one of possession, whereas the relationship in the other statements has to do with friendship.

### 2.1. *Additional Coding Choices for Map Analysis*

There are a large number of choices that must be made by the researcher when doing map analysis. As map analysis subsumes content analysis, all of the issues previously raised still apply and the same set of choices must be made. In addition, choices also arise concerning the possible concept types, the relationships between concepts, and the use of social knowledge. As with content analysis, different tools for doing map analysis may place different restrictions on the researcher as to which of these choices are available. For example, some procedures predefine a set of concept types or the meaning of relationships. Others, such as MECA, do not.[11]

#### 2.1.1. *Concept Types*

The basic issue here is how many categories or types of concepts should be used. The researcher needs to choose how many types of concepts are to be used. In addition, if more than one type is needed, the researcher must develop a set of concept types that span the set of categories of interest in that research. Despite the work by numerous researchers in multiple fields, no single overarching set of concept types has emerged that is valuable across all research questions. In many studies, one category of concepts is sufficient. The only justification for multiple categories of concepts occurs if there are predefined classes of concepts, each of which the researcher wants to examine separately. If there are no predefined categories then only one category of concepts should be used and concepts should be classified into categories after the coding.

The advantage of a single concept type is that it facilitates rapid and more automated coding. This is particularly important for exploratory research or when the researcher wants the data to dictate what is coded. Palmquist (1990) examined the difference in learning in two classrooms in which "research writing" was taught. Because the study was highly exploratory, all concepts were treated as though they were in a single category. Using multiple concept types, however, can facilitate analysis as it allows the researcher to examine

[11]In order to use the MECA procedures for coding maps, the researcher must first use STARTUP, the MECA routine for storing the set of coding choices being made. See appendix for additional details. See also Carley and Palmquist (1992) for an extended example.

differences between categories of knowledge. This is usually more appropriate when the researcher has either predetermined concepts, or has a theoretical interest that specifies different claims for different types of knowledge. For example, Carley (1984, 1986) used map-analytic techniques to examine differences in students' views of what they wanted in a tutor. In this study, four categories of concepts were used because of the different roles that the types of knowledge were thought to play in decision making: aspects of the tutoring job, requirements of the tutoring job, facts about candidates, and qualities of candidates. As another example, Cicourel and Carley (1990) used map-analytic tools to examine the role of social knowledge in children's memory of stories. They utilized six concept types corresponding to story components: people, places, actions, objects, adjectives, and statements. They note that categorizing concepts into these types is a de facto recognition by the researcher of knowledge that is social—that is, shared by the researcher, the story author, and the readers of the story.

2.1.2. *Level of Analysis—What Constitutes a Relationship*

When a map-analytic approach is taken, the researcher can choose how much information to preserve about the relationship. The researcher may choose to simply note that two concepts are related. This choice implies a specific series of secondary choices—all relationships are of the same strength, all have the same sign, all are bidirectional, and all have the same meaning (i.e., differences in meaning are not preserved). This choice is appropriate when the researcher is engaged in exploratory research or when there are no theoretical reasons to distinguish between types of relationships. For example, Thomas (1991) examined shifts in Americans' perception of "robots" as portrayed in American science fiction literature in the past 50 years. Because this study was highly exploratory and because only a few texts were being analyzed, Thomas used relationships to simply denote the existence of a link between two concepts. Similarly, Palmquist (1990) also used existence relationships. Alternatively, the researcher may choose to preserve a large amount of information about each relationship, including its strength, sign, direction, and meaning. Each of these secondary choices will now be discussed.

1. *Strength.* The issue is whether relationships should vary in strength. When all relationships are the same strength, then maps are easier to compare, and many graph comparison techniques are applicable. Having multiple strengths, however, lets the researcher retain more information and makes the resultant maps a closer replica of the underlying text. Such an approach is especially useful when the researcher wants to use the coded maps to augment more qualitative discussions. Having strengths is also useful when the researcher is interested in the emphasis that is placed on the relationship between concepts. For example, Carley (1986) used a range of 1 to 3 to denote whether the relationship was implied in the text (1), stated explicitly (2), or emphasized (3). Strength can also be used to denote frequency. For example, Danowski (1982) conducts map analyses using proximity and defines the strength of the relationship as simply a count of the number of times the two concepts are proximal within the text.
2. *Sign.* Sign is related to whether the researcher can identify a positive or negative relationship between the two concepts. Consider the statements "Jess fits in" and "Frank does not fit in." The first statement could be coded as a positive link between "Jess" and "fits in." The second statement could be coded as a negative link between "Frank" and "fits in." Or, the second statement could be coded as a positive link between the concept "Frank" and the concept "does not fit in." The difference here is in the location of the negation. In the first example, the negation is placed on the relationship, thus implying that "fits in" and "does not fit in" are conceptual opposites. In the second case, negation is directly attached to the concept implying that the concepts "fits in" and "does not fit in" are simply different concepts. Which is the more appropriate coding scheme for texts is a point for future research. Placing negation on the concepts facilitates map comparison, however. For example, Carley (1984; 1986) used both positive and negative relationships and so had to define procedures for comparing statements with the same concepts but relationships of different signs. In contrast, Palmquist (1990), Cicourel and Carley (1990), and Thomas (1991) used only positive relationships and, where necessary, used "negative"

concepts. As a result, it was easy to combine and compare maps to examine differences across texts. As a final note, in some cases sign is not an issue. For example, Carley and Kaufer (forthcoming) coded general social knowledge about "Comedy" and "Drama" as available in a thesaurus. Since only synonyms were coded, negation was not an issue; that is, in this case all relationships can be thought to denote the positive sentiment "is a synonym of," and since only synonyms were used there was no need to denote the negative "is not a synonym of."

3. *Direction.* Direction has to do with whether the first concept is seen to have some type of "prior" relationship to the second concept. Various types of prior relationship can be thought of—for example, "*a* implies *b*," "*a* comes before *b*," "if *a* is true, then *b* is true," and "*a* qualifies *b*." Choosing to ignore directionality (or equivalently to treat all relationships as bidirectional) is appropriate when the research is exploratory and facilitates automatic coding. When a proximity approach is taken directionality is often ignored. For example, Palmquist (1990) treated all relationships as bidirectional. In contrast, coding directionality can provide information about the way in which the impact of new information propagates through the network and affects decisions (Carley 1984), and the structure of meaning. If a researcher is interested in separating procedural or temporal knowledge from declarative or definitional knowledge, eliminating directionality can make it impossible to locate these differences.
4. *Meaning.* Relationships can also vary in meaning. In many studies, one type of relationship (one category of "meaning") is sufficient as the dominant interest of the researcher is simply to determine which concepts are related. For example, when a proximity approach is taken, all relationships are of the type "is proximal to." In fact, in many of the studies previously mentioned (Thomas 1991; Palmquist 1990; Saburi-Haghighi 1991; Carley and Kaufer, forthcoming), there was no need for preserving any aspect of the relationship other than its strength, direction, and sign. Nevertheless, preservation of meaning may be especially useful in story understanding and in examining semantic structures where it is important to distinguish among temporal, logical, and definitional linkages. In these cases the researcher must develop a set of relationship types that span the set of "meanings" of interest in that

research. This is analogous to the selection of concept types by the researcher. And, as with concept types, despite the work by numerous researchers in multiple fields, no single overarching set of relationship types or meanings has emerged that is valuable across all research questions. For example, Carley (1988) defined three types of relationships—definitives, logicals, and simples—in order to use the expert's knowledge of the sociolinguistic environment to automatically explicate information implicit in the text. As another example, Cicourel and Carley (1990) were interested in distinguishing those statements that were spontaneously provided by the subjects and those that occurred only after prompting. They thus employed two types of relationships: "normal" and "after prompting" (Cicourel and Carley 1990).

2.1.3. *Level of Implication for Relationships*

The researcher also needs to determine the desired level of implication for relationships. For example, are relationships to be coded only when explicit words occur in the text linking two concepts (as "is" links "Liesel" and "boy's mother" in the statement "Liesel is the boy's mother")? Or are relationships to be coded even when the link is implied (as in the statement "Liesel, the boy's mother")? Coding only explicit relationships facilitates automatic coding and may minimize errors due to the coder "reading in" information into the text. In contrast, coding implicit relationships provides a richer, more detailed map and facilitates cross text comparison.

2.1.4. *Social Knowledge*

The researcher needs to determine whether social knowledge is to be included in the coded texts. Social knowledge can be defined as knowledge that one can reasonably expect all members of the sociolinguistic environment to share—for example, that parents are older than their children.[12] Clearly, the determination of what knowledge is social knowledge varies across cultures and time. Vast quantities of social information are implicit within any text, regardless of the source—interviews, children's stories, or newspaper articles. Different members of the same sociolinguistic environment may have different proclivities for explicating such knowledge in the texts they

[12]For more extensive discussions of social knowledge, see Carley (forthcoming b).

generate. Thus adding social knowledge can actually bring out the hidden similarities in texts. Consider the following two passages:

> Joe's a gnerd who always studies in the library.

> Joe's a gnerd who always studies in the library and doesn't fit in. His door is never open, he works too hard.

Coding only the explicit knowledge leads to maps that appear very different, as illustrated in Figure 1.[13]

For ease of illustration, this coding was done using the following choices: only one type of concept used; relationships can differ in direction and sign; and all relationships are of the same type and same strength. Furthermore, relationships are coded based on whether there is an implicit semantic connection between the concepts. When a semantic approach is taken, each statement forms a simple sentential unit—such as "if *a* then *b*" or "*a* is an example of *b*" or "*a* ‹verb› *b*."[14] Since all relationships are treated as being of the same type, the specific semantic connection between the two concepts is not preserved. These particular coding choices tend to generalize relationships and emphasize the degree of similarity between the texts. Because these particular coding choices will be used in later illustrations in this paper, this set of choices will be referred to as the "simple semantic method."

As Carley argued (1984, 1988), within the sociolinguistic environment that produced these texts there is a set of common understandings or social knowledge that the students tend to take for granted. These include (but are not limited to) the following:[15]

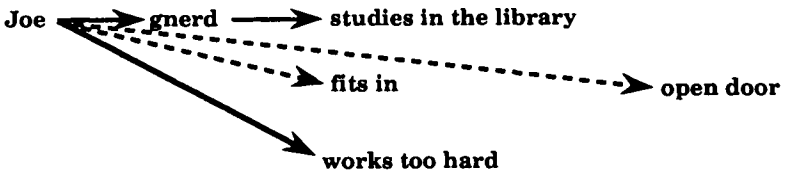
[13]These texts were first coded using CODEMAP, which produced a data base for each text. The actual drawings were produced using a two-step process. First, the data bases for the maps were run through the MECA program DRAWMAP, which generates a MacDraw data base file. These files were then ported to MacDraw, where they were annotated to produce the figures seen here. See appendix for additional details.

[14]For additional information on two different procedures for coding relationships from a semantic perspective, see Cicourel and Carley (1990).

[15]Each of these statements is, in Carley's (1988) terms, a definitive and so indicates an implication that is agreed upon by all individuals in the group. They have been written here as English statements for ease of understanding, rather than in the format used by Carley.

**MAP 1: Joe's a gnerd who always studies in the library.**


**MAP 2: Joe's a gnerd who always studies in the library and doesn't fit in. His door is never open, he works too hard.**




| | |
|---|---|
| **Shared Concepts** | **3** |
| **Shared Statements (1 bidirectional= 2 relations)** | **2** |
| **Shared Concepts given Shared Relationships** | **2** |
| **Concepts Map-1 Only** | **0** |
| **Concepts Map-2 Only** | **3** |
| **Statements Map-1 Only** | **0** |
| **Statements Map-2 Only** | **3** |

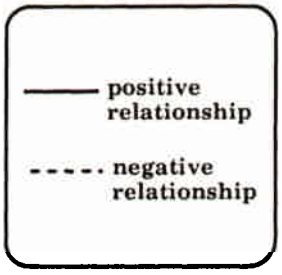



**FIGURE 1.** Maps coded using only explicit information.

- If someone studies in the library, then his/her door is never open.
- If someone studies in the library, then he/she is not accessible.
- If someone has an open door, he/she interacts with students.
- If someone is a gnerd, then he/she works too hard.
- If someone is a gnerd, he/she is not friendly.
- If someone is a gnerd, he/she interacts with students.

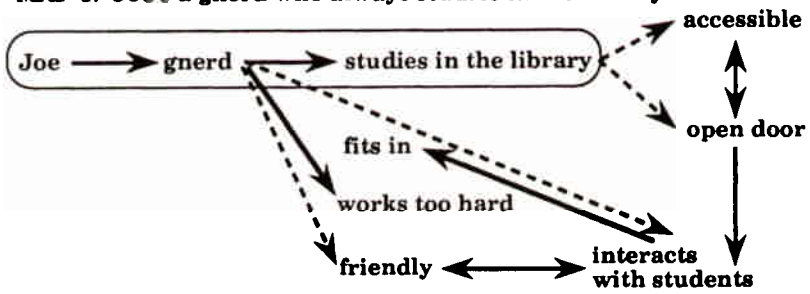
- If someone interacts with students, he/she fits in.
- If someone interacts with students, he/she is friendly.
- If someone is friendly, he/she interacts with students.
- If door is open, then he/she is accessible.
- If he/she is accessible, then door is open.

This social knowledge lies behind the texts. Making this implicit information explicit demonstrates that there is a great deal of shared meaning between the two texts (Figure 2).[16] As with Figure 1, a simple semantic method was used to code the maps in Figure 2. For further discussion of a procedure for making such implicit knowledge explicit, see Carley 1988.

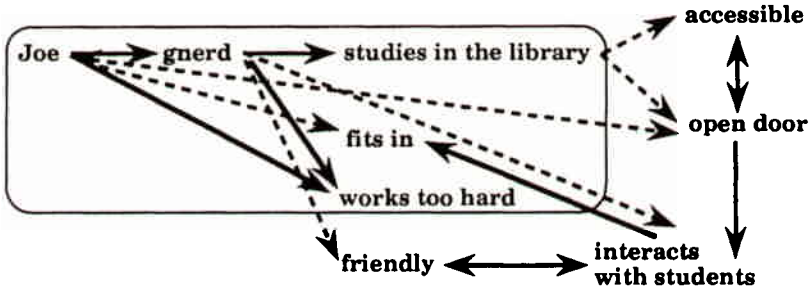
Explicating maps using social knowledge can, to an extent, overcome underestimations of similarity in texts. Furthermore, such explication can minimize the need to extensively train coders, thus allowing the researcher to utilize more "novice" labor, and it can increase the reliability and validity of the coded maps (Carley 1988). Nevertheless, such a procedure is not a panacea. Misestimation of social knowledge can make texts appear more similar than they truly are. In addition, the researcher must be careful not to be overly normative in making judgments about social knowledge. Comparison of verbal discourse strategies with written texts from the same speech community can be helpful here. In addition, there are many research questions where utilization of such a procedure would be inappropriate. Such a case may occur if the researcher is interested in comparing a "naive" respondent's viewpoint and that of an expert, which means that the utilization of expert knowledge to explicate maps is incorrect. For example, when Palmquist (1990) compared students' conceptions of research writing with the teacher's conception, explication of the students' maps with the expert's (in this case the teacher's) knowledge would have been inappropriate. A second such case occurs if the researcher is interested in comparing how different individuals portray their information. In this case the correct analysis uses information explicit in the text and not the underlying social knowledge.

[16]These modified or explicated maps were produced using the MECA program SKI. See the appendix for additional details.

**MAP 1: Joe's a gnerd who always studies in the library.**




**MAP 2: Joe's a gnerd who always studies in the library and doesn't fit in. His door is never open, he works too hard.**




| | |
|---|---|
| **Shared Concepts** | **9** |
| **Shared Statements (1 bidirectional= 2 relations)** | **13** |
| **Shared Concepts given Shared Relationships** | **9** |
| **Concepts Map-1 Only** | **0** |
| **Concepts Map-2 Only** | **0** |
| **Statements Map-1 Only** | **0** |
| **Statements Map-2 Only** | **3** |

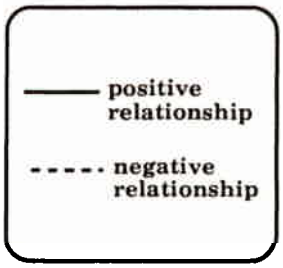



**FIGURE 2.** Maps recoded using implicit social knowledge.

### 2.2. *Advantages and Limitations*

There are several advantages to map-analytic techniques: (1) They allow the researcher to examine the micro-level differences in individuals' maps and to get at differences in meaning. (2) They subsume content-analytic techniques; thus, if the researcher starts by coding maps, it is always possible to backtrack and do a more traditional content analysis. (3) They facilitate examining hierarchies of meaning, as happens when categorizing concepts or types or relationships. (4) They typically provide researchers with the ability to stay close to the text but also to move beyond it as they are given the luxury of alternating between actually examining what specific concepts and relationships occur in a text and statistically analyzing the coded form of the texts. In addition, since all knowledge is coded in terms of statements, it is in many cases possible to interpret these statements (if the relationships are unidirectional) as rules, and thus as the knowledge base for an expert system. This in turn facilitates examining the decision-making process implicit within the map. A final advantage is that by focusing on the structure of relations between concepts and relations, and the relationships between concepts, the attention of the researcher is directed toward thinking about "what am I really assuming in choosing this coding scheme?" Consequently, researchers may be more aware of the role that their assumptions are playing in the analysis and the extent to which they want to, and do, rely on social knowledge.

A disadvantage to map-analytic techniques is that they are harder to automate. In particular, it is harder to automate the coding of relationships than it is to automate the coding of concepts. Coding relationships appears to require a higher level of interpretation than the coding of concepts. If this is the case, then analyses based on maps rather than just concepts may be more prone to error. Whether this is true, however, is a point for future research.

## 3. COMPARISON OF ANALYSIS TECHNIQUES

In this section various additional comparisons are made in order to further illustrate how different coding choices affect the results of textual analyses. First, content analysis and map analysis are contrasted, thus illustrating the effect of choosing to examine only con-

cepts or concepts and relations. Second, multiple map-analytic approaches, each of which uses a different procedure for defining the relationship between concepts—semantic-based, proximity-based, and story-line–based—are contrasted. Finally, concordance analysis is examined in order to illustrate the effect of using a reduced form of the text for doing either content analysis or map analysis.

### 3.1. *Content Analysis and Map Analysis*

Both content-analytic and map-analytic techniques can be used on the same texts. In such cases, as was suggested earlier, each technique may lead to a different interpretation of the texts. The extent and nature of these differences are going to depend on the coding decisions made by the researcher, as in the example cited earlier, where the texts of Students A and B, which discuss scientists, were coded from a content-analytic perspective using only explicit concepts. This demonstrated that two texts, though differing in meaning, could appear identical when only concepts were coded. These two texts were recoded as maps, using the simple semantic method previously described.[17] The resultant maps appear at the top of Figure 3 and statistical information on the two maps appears at the bottom of this figure.[18] Note that when relationships are ignored, the information for shared concepts is identical to the information garnered via traditional content analysis (refer back to Table 1).

In contrast to the content-analytic case, when a map-analytic approach is taken, the two texts appear quite different. In fact, although they still share all the same concepts, they share only five of the statements. When we consider only these five shared statements, we find that the two students have only five concepts in common. In other words, although they use the same words, they use only five of the concepts in anything approaching the same way. Specifically, Student A reports on what "I" found about "scientists" and has an elaborated notion of "scientists" as doing "research." Thus "research" is needed to make "discoveries." In contrast, Student B focuses on the "research" conducted by "I" and does not see "scientists" as doing "re-

[17]The MECA procedure CODEMAP was used to code the text as a data base from which the figure was produced using DRAWMAP.

[18]Statistical information was collected using the MECA program COMPRA. See appendix for details.

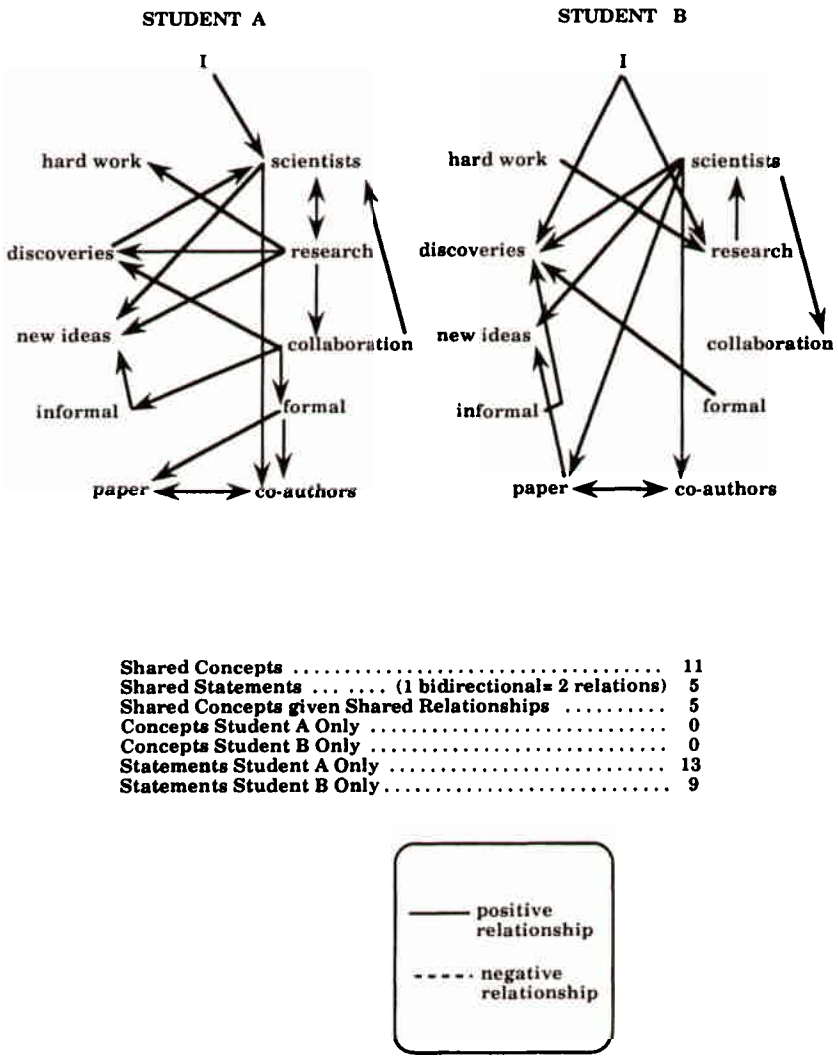



**FIGURE 3.** Coded maps for comparison with content analysis.

search" but as simply making "discoveries." For Student A, the key theoretical constructs (most lines in and out of the concept node) are "scientists," "research," and "collaboration"; whereas for Student B, the key constructs are "scientists" and "discoveries."

### 3.2. *Proximity Analysis*

When a map-analytic perspective is taken, relationships can be coded from a variety of approaches other than the semantic approach described above. For example, the researcher might employ a proximal (Danowski 1982) or temporal (Roberts 1989; Dyer 1983, for example) approach. When a proximity approach is taken, a relationship is placed between two concepts in the event that they occur within some prespecified window (for example, within the same sentence). While using proximity to define relationships greatly facilitates automatic coding, it can obscure meaning. Consider the following two passages:

> The president fired James and promoted Karin.

> The president fired Karin and promoted James.

Now assume that a preparser is being used such that all articles and conjunctions are deleted.[19] Then, using the sentence as the "window" length, a proximity approach would locate maps like those in Figure 4. Here, since all the same words occur within each sentence, the maps for both sentences are identical. Moreover, the proximity approach cannot automatically distinguish the sign of a relationship, nor the direction (other than by word order), nor the meaning. The strength of the relationship can be, but need not be, measured by the number of times the two concepts are proximal. When a less automated approach like that proposed by Carley (1986; 1988; 1990*a*) is used, distinctions between relationships can be drawn. For example, these two texts were coded using the simple semantic method and the resultant maps are displayed in Figure 5. Under this coding procedure, the maps for both sentences are different, with directionality indicating who did the firing/promoting and who was fired/promoted and with the sign showing that firing and promoting are in some sense "opposite" relationships.

[19]Within the MECA package, DELCON could be used for this purpose.

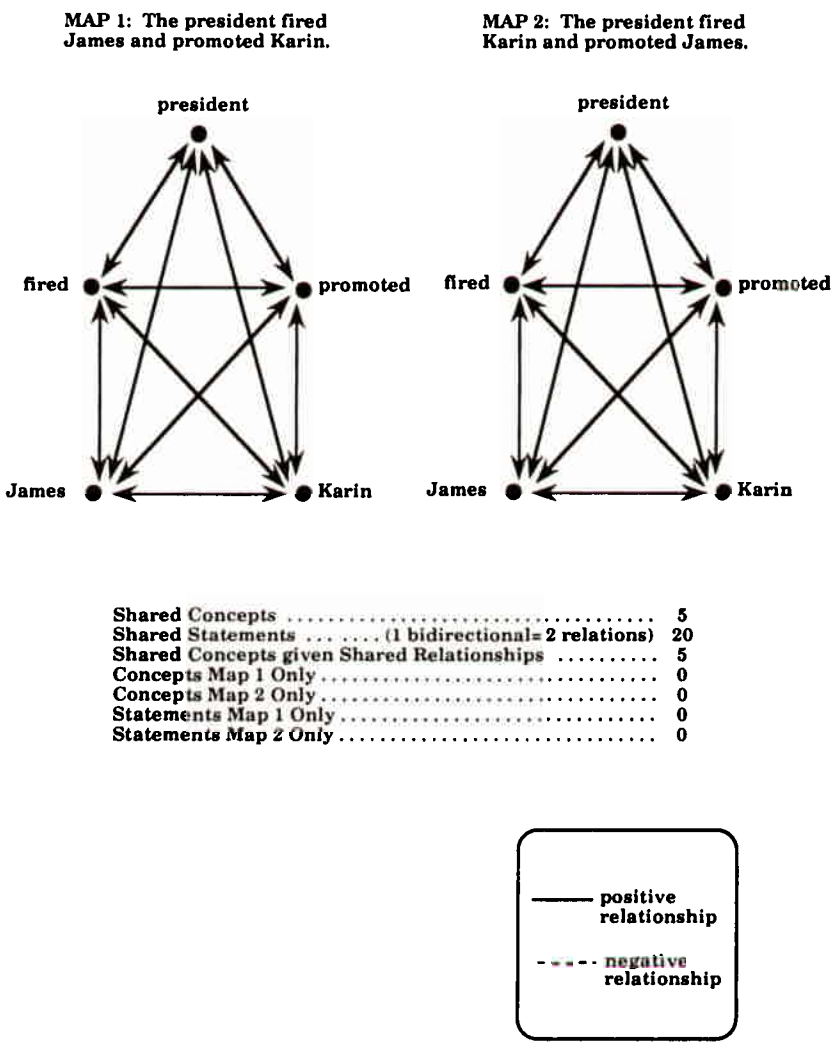



**FIGURE 4.** Maps produced by proximity analysis.

### 3.3. *Linguistic Approaches*

Now let us contrast map-analytic techniques that take a semantic approach with more linguistic, story-line, approaches that, though concerned with semantics, emphasize syntax—linguistic content analysis (Lindkvist 1981; Hutchins 1982; Roberts 1989), semantic

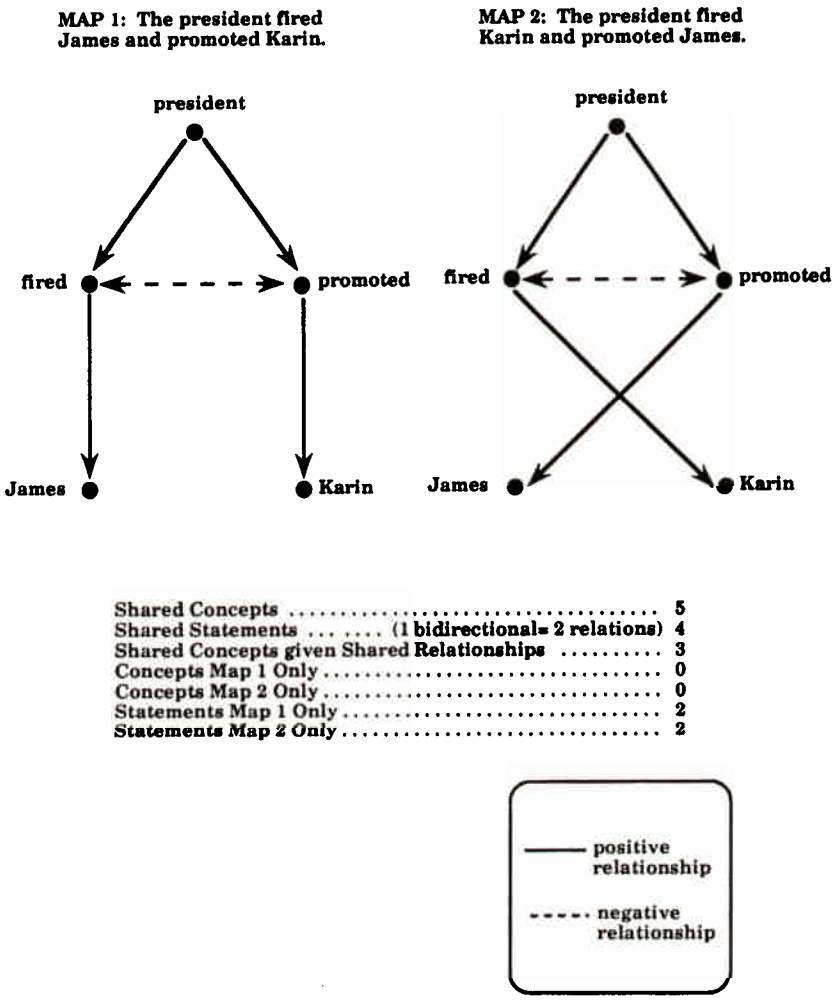



**FIGURE 5.** Maps produced using CODEMAP.

grammars (Franzosi 1990*a*, *b*), and story processing (Lehnert and Vine 1987; Dyer 1983, for example). All such approaches produce networks of concepts or maps and can be thought of as map-analytic approaches. The major difference between these approaches and the purely semantic approach described above is one of emphasis. Map-analyses that take a semantic approach emphasize the conceptual or

definitional relationships between concepts, whereas a story-line approach would emphasize the sequential story relationships.

This difference can be seen by contrasting the way these two approaches would code the sentence "Cassi throws a ball to Corwin." A story-line approach always, regardless of the focus of the research, would code this sentence in a form like—‹sending-actor› Cassi ‹action› throws to ‹object› ball ‹object-modifier› a ‹receiving-actor› Corwin. In terms of the map representation, there are in this example five types of concepts: sending-actor, action, object, object-modifier, and receiving-actor. In contrast, the coding resulting from a purely semantic approach depends on the researcher's focus. If the focus is on locating definitions of "ball," then only a single type of concept might be used and only a single statement coded containing the concepts "ball" and "throw" and a relation from ball to throw. Whereas, if the focus is on examining Cassi's actions, then the researcher might use two types of concepts, "actors" and "actions," and code two statements, the first with a relationship from "actor:Cassi" to "action:throw ball" and the second with a relationship from "action:throw ball" to "actor:Corwin."

An advantage of story-line approaches is that the researcher can make use of shared sociocultural knowledge about the appropriate relationship between concept types to minimize coding effort. Such a facility is not, at least currently, a part of the purely semantic-based or proximity-based approaches. A further difference is that semantic-based and proximity-based approaches lend themselves to, and are typically used for, generating graphs and numerical evaluations of the text, while story-line techniques lend themselves to processes for regenerating the story given the coded representation. Semantic and proximity approaches are appropriate when examining decision processes, mental models, definitions, conceptual frameworks, and the role of social knowledge within these. Story-line approaches are appropriate when examining the sequence of actions and the similarity between different sequences of action.

### 3.4. *Concordance Analysis*

A concordance is an alphabetical list of the words (or concepts) in a text with reference to the passage(s) in which they occur. It is typi-

cally used in examining plays, poetry, songs, and religious texts such as the Bible. Analytic concordances match words in one text with the corresponding word in a translated version of the text—for example, words in the English Bible may be matched with their Hebrew original (Young et al. 1936). Lexical concordances try to classify words by their signification, as in Ellis's (1968) concordance of Shelley's poetry. Classification schemes, however, vary across concordances. Advances in automated textual analysis and artificial intelligence are making it possible to, under certain conditions, automatically generate a concordance and associated lexical statistics that enable content analysis and facilitate map analysis (such as the frequency with which words occur). The main advantage of a concordance to qualitative research is that information on where the word occurs in the original text is preserved. Semiautomated procedures such as ASKSAM retain this feature. Although concordances have not typically been used to perform content or map analysis, they can be used in either procedure. For example, researchers can use a concordance to do content analysis if they wish to make use of the exact words or phrases in the text and are willing to either do without implication, translation, and generalization or to develop a thesaurus based on the words in the concordance that automates this process.

To illustrate this type of analysis, the statements of Students A and B will be used and each sentence will be treated as a separate "passage." The resulting concordances are displayed in Table 2. As can be seen, Student A uses 64 words and Student B uses 48 words. And, although Student A and B both use "scientists" four times, A uses it three times in one sentence (sentence 2), whereas B uses it in all three sentences. As noted, a concordance can be used to do either a content analysis or a proximity analysis. Given the set of concepts used in Table 1, one can extract from the concordance the frequency with which each of those concepts was used, thus facilitating content analysis. Alternatively, a concordance can be sorted by location and then, given a set of concepts, a map can be drawn. In this case, the relationship would be based purely on proximity. The strength of the relationship could be characterized as the number of times the two concepts co-occurred. Based on the set of concepts in Table 1, the resultant maps coded from the concordance using proximity are shown in Figure 6. For ease of comparison with other codings of

TABLE 2
Concordances for Two Texts

| Student A | | | Student B | | |
|---|---|---|---|---|---|
| Concept | Number | Sentence | Concept | Number | Sentence |
| I | 1 | 1 | I | 1 | 1 |
| a | 1 | 3 | It | 1 | 1 |
| and | 2 | 1 2 | My | 1 | 2 |
| are | 1 | 3 | Some | 1 | 3 |
| as | 2 | 3 3 | and | 2 | 1 3 |
| be | 1 | 3 | are | 1 | 2 |
| by | 1 | 2 | at | 1 | 2 |
| coauthors | 1 | 3 | coauthors | 1 | 2 |
| collaboration | 2 | 2 3 | collaboration | 2 | 1 2 |
| discoveries | 2 | 1 2 | containing | 1 | 2 |
| engage | 1 | 1 | discoveries | 2 | 1 3 |
| famous | 1 | 2 | engaged | 2 | 1 2 |
| formal | 1 | 3 | famous | 1 | 1 |
| found | 1 | 1 | formal | 1 | 3 |
| generate | 1 | 1 | hard work | 1 | 1 |
| hard work | 1 | 2 | have | 1 | 3 |
| in | 2 | 1 1 | in | 2 | 1 2 |
| informal | 1 | 3 | informal | 1 | 1 |
| involves | 1 | 2 | least | 1 | 2 |
| is | 1 | 2 | made | 1 | 1 |
| leads | 1 | 2 | make | 1 | 3 |
| lunch | 1 | 3 | many | 1 | 1 |
| make | 2 | 1 2 | new ideas | 2 | 2 3 |
| may | 1 | 3 | of | 1 | 2 |
| new ideas | 2 | 1 3 | one | 1 | 2 |
| of | 1 | 3 | other | 1 | 2 |
| often | 1 | 2 | paper | 1 | 2 |
| or | 1 | 3 | research | 2 | 1 2 |
| order | 1 | 1 | scientists | 4 | 1 2 2 3 |
| other | 1 | 2 | showed | 1 | 2 |
| over | 1 | 3 | that | 1 | 2 |
| paper | 1 | 3 | their | 1 | 2 |
| research | 2 | 1 2 | to | 1 | 1 |
| scientists | 4 | 1 2 2 2 | was | 1 | 1 |
| share | 1 | 3 | with | 1 | 2 |
| such | 4 | 2 3 3 3 | | | |
| that | 1 | 1 | | | |
| the | 1 | 2 | | | |
| they | 2 | 3 3 | | | |
| to | 2 | 1 2 | | | |
| when | 2 | 3 3 | | | |
| which | 2 | 2 2 | | | |
| with | 1 | 2 | | | |

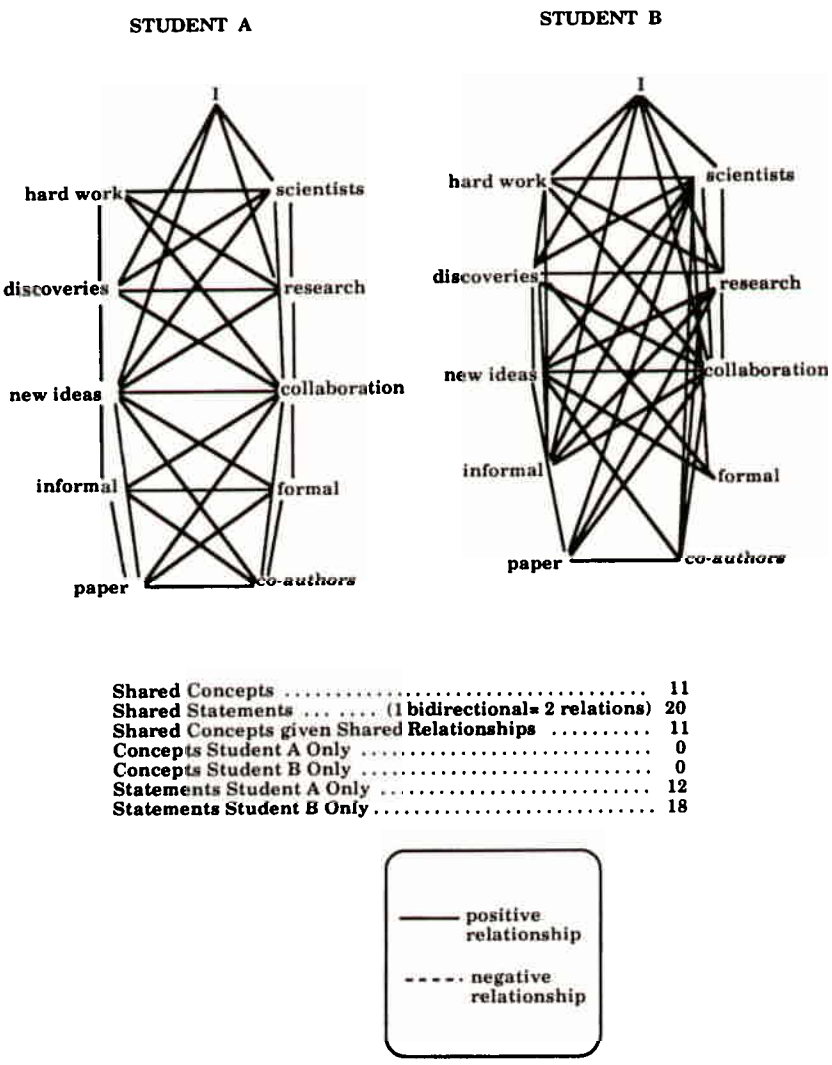



**FIGURE 6.** Coded maps based on concordance.

these texts, the strength of all relationships is identical; thus strength simply denotes whether there is or is not a relationship between those concepts. As can be seen, this proximity based coding overestimates the similarity of the texts compared with that produced when a more semantic approach is taken, as in Figure 3.

## 4. THE NEXT STEP

Once frequency counts of concepts and sets of statements have been extracted from the texts, the process of textual comparison begins. A detailed discussion of how to compare texts is well beyond the scope of this paper. Nevertheless, in keeping with the foregoing discussion, it is worth pointing out that there is a series of analysis choices that goes beyond specific coding and analysis procedures. Ideally, the tools that help the researcher to code the texts put the resultant information in a data-base form that facilitates comparison, regardless of the choices made by the researcher. Unfortunately, this is not always the case. Thus knowing what choices need to be made and the types of factors that should be considered should aid the researcher in selecting tools.

When content-analytic techniques are used, text comparison is fairly straightforward and typically focuses on the number of concepts shared and whether there is a pattern to what concepts are shared or not shared. Standard procedures for categorical analysis, clustering, grouping, and scaling are all useful. Map-analytic techniques provide the researcher with a greater number of choices. For example, maps can be compared either in terms of shared concepts, shared statements, or both. Comparison of maps in terms of shared concepts moves the researcher back into the realm of content analysis and facilitates comparison of map-analytic tools with content-analytic tools. Comparison of maps in terms of statements facilitates locating differences in meaning. Using measures of similarity for both concepts and statements is often the most satisfying as it enables the researcher to discuss the extent to which the authors of the texts have the same words but different meanings. For example, when comparing two maps in terms of concepts, the researcher can choose to compare the maps in terms of what concepts are or are not present or to compare maps in terms of what concepts are or are not present given the statements that the two maps have in common. Contrasting these two comparisons provides insight into differences in meaning across the texts. For example, two maps may contain all the same concepts (e.g., as in Figure 3), but when the statements are considered they share only five of the eleven concepts. In other words, although they use many of the same concepts, they use them in different ways.

The ease with which maps can be compared in terms of statements depends on how the statements are coded and on the number of types of concepts. Indeed, what coding choices are made affects the range of tools available for analyzing and comparing the resultant map. For example, if all concepts are the same type, if the relationships are identical in terms of sign, strength, and meaning, and if the relationships are all bidirectional, then standard graph analysis techniques can be used. In contrast, simply making the relationships directional moves the researcher into the realm of digraph analysis. Whereas having multiple types of concepts moves researchers into the realm of colored graphs, and having multiple strengths on the relationships moves them into the realm of network analysis.

A related issue is the level of similarity between statements that is necessary to say that two maps contain the same statement. This issue depends in part on how relationships are coded. For example, assuming that relationships are simply coded as either existing or not existing, that there are no differences in strength, sign, directionality, or meaning, and that the two maps have a relationship between the same two concepts, then the two maps can be said to contain the same statement. This approach was used by Palmquist (1990), Thomas (1991), and Saburi-Haghighi (1991). More complex schemes for coding relationships require that the researcher define schemes for combining statements and making similarity judgments (for example, see Carley 1984). Consider again Figure 3. In this case relationships are allowed to vary in directionality. As a result, the two maps are found to share five statements. If all relationships had been treated as bidirectional, then three additional statements would be shared (hard work—research; discoveries—scientists; and scientists—collaboration). Treating a statement with a bidirectional relationship as two statements would lead to the following statistics in place of the ones at the bottom of Figure 3:

- Shared Concepts 11
- Shared Statements 14
- Shared Concepts Given Shared Relationships 8
- Concepts Student-A Only 0
- Concepts Student-B Only 0
- Statements Student-A Only 20
- Statements Student-B Only 12

Another important consideration is that even when social knowledge is included in the coded maps, it tends to be quite sparse. That is, there are very few relationships per concept. For example, for Carley (1984, 1986) the average number of statements per concept per map was four. Although the distribution of statements across concepts is very nonuniform, with some concepts occurring in only one statement and others in over 20 statements. This sparsity affects what analysis techniques are viable and also the storage cost of various database schemes. One area where this is significant is when one wants to contrast actual statement usage with potential usage. Barring linguistic constraints on what types of concepts can be related to what other types, in general, there are potentially between each pair of concepts as many relationships as four times the number of possible strengths times the number of possible meanings. This assumes that there can be one relationship for each direction and for each sign. Assuming even a highly simple coding scheme in which relationships simply denote existence (which means there is no differentiation in strength, sign, directionality, or meaning), there are $N(N - 1)$ possible statements (assuming concepts cannot be related to themselves). Thus, for even a moderate-sized data base of about 200 concepts, the number of possible statements would be 39,800, but the number of statements in any one map might be more on the order of 100.

A further choice is whether maps are to be compared quantitatively or graphically (or both). One of the advantages of map-analytic techniques is that the researcher can use both modes of analysis in a complementary manner. Quantitatively comparing maps allows for a basic understanding of the degree of similarity and difference. Visual comparison of the graphs helps motivate rich detailed discussions of the factors in the individual and the sociolinguistic environment that led to these differences and enhances qualitative studies. In general, choices about how to compare maps can be made, unlike the other choices discussed, after the researcher has coded the maps.

This brief discussion does not begin to cover the full extent of the range of issues present in analyzing the coded textual data. Furthermore, this discussion should not be taken as evidence that there exists an analysis procedure for any coding scheme, or that all analysis issues have been solved. In particular, the greater the complexity in coding relationships the lower the likelihood that an appropriate

analysis tool exists. In general, our ability to code texts and retain vast information on both concept types and relations far exceeds our ability to analyze the resultant data. A great deal of additional research is needed on exactly how to combine and compare maps under more complex coding scenarios.

## 5. FINAL REMARKS

Both content-analytic and map-analytic techniques help the researcher to code and analyze texts. The two approaches can be carried out manually or with the aid of computer-based tools. For effective use of such tools, the researcher will need to know the specific procedures followed by the coding software, an issue that is beyond the scope of this paper. There are, as was detailed here, a set of coding choices and some general guidelines that transcend particular coding protocols. These choices, the eight that are needed whenever concepts are coded and the additional four that are needed to engage in map analysis, should in general be made prior to selecting a coding procedure (manual, computer-assisted, or completely automated). These coding choices affect the conclusions that the researcher can draw from the data. Associated with these choices are a set of guidelines. The guidelines presented here are garnered from a variety of experiences in coding texts using both content-analytic and map-analytic procedures. These experiences suggest that the researcher should make choices that simplify the coding scheme—e.g., reduce the number of concepts (for both content and map analysis) and the number of concept and relationship types (for map analysis) to a manageable size that admits a meaningful comparison of texts.

Using the simplest possible coding scheme given the research question will increase the ease of coding, the ease with which texts can be compared, the extent to which the comparisons are meaningful, and the likelihood that existing tools can be used for automating the coding and analyzing the resultant coded data. Nevertheless, a simple coding scheme will lock the researcher into a certain way of analyzing the data. Data coded via more complex schemes can always be recoded during analysis to a more simple scheme. The ease with which such recoding is possible, however, depends on the particular coding procedures employed.

Regardless of the procedures employed, textual analysis is very

detailed, time-consuming, and tedious work. When coding texts, people are quite prone to making errors. These errors can occur for a variety of reasons, such as inconsistency in generalizing concepts and changing what information is considered relevant part way through and then failing to recode earlier texts. There are several advantages to computer-based tools that either completely automate the coding process or simply assist the researcher. These include, but are not limited to, the following: They decrease the amount of time required to code texts; they eliminate or minimize errors due to inconsistency; they eliminate errors due to miscounting; they often eliminate spelling errors; they encourage systematicity; and they put the data into a form where it is easier to compare analytically. Furthermore, the startup costs of using computer-based tools are substantial enough that they encourage researchers to spend time thinking through coding choices such as those discussed here. Such tools do not, however, make textual analysis trivial. Further advances, particularly in map-analytic techniques, are needed to create tools that more completely automate the coding process and make it easier for the researcher to dynamically change coding choices and then compare the results.

Computer-based tools are most helpful when the researcher is interested in analyzing a large number of texts or texts that are extremely lengthy. Even for as few as ten texts, these approaches are very valuable. This is due in part to the possible reduction in errors. In addition, in the case of the map-analytic techniques, once the maps are coded it is relatively easy to extract vast quantities of empirical information. Such data make feasible analyses that were heretofore extremely difficult if not impossible. Map-analytic techniques focus on the extraction of both concepts and the relationships between them. Since this subsumes content analysis, the results can also be analyzed from a traditional content-analytic perspective. The addition of relational information, however, allows the researcher to examine not just shifts or differences in word usage across time and/or people but also shifts or differences in meaning.

## APPENDIX: EXAMPLE OF CODING, DISPLAYING, AND ANALYZING MAPS

This appendix provides a brief overview of the various MECA routines for coding and analyzing texts. Software routines, other than

those in MECA, exist for doing content analysis and map analysis, but they will not be discussed. Additional information on the MECA procedures for coding maps is available in Carley (1988, 1990*a*) and Carley and Palmquist (1992). Many additional details of the programs that are not particularly germane here, such as storage requirements and exact file formats for input and output files, are suppressed. See Carley (1990*a*) for details.

### *Content Analysis*

Three procedures exist within the MECA package that aid the researcher in doing content analyses: DELCON, TRANSLATE, and CMATRIX2.

DELCON: A program for deleting concepts from a text.

*Input:* 1. A file containing the list of concepts to be deleted. This list might contain the concepts—Killishandra, Ezra, he, she, them, and, or, the, in, order, and to—as well as notation such as ",", "'", and ".".

2. A file containing the list of texts that the researcher wants to delete information from.

3. A set of files, one for each text, from which the researcher wants to delete the irrelevant concepts. One such text, "tailed," might contain the passage described earlier about Killishandra and Ezra tailing the merchant.

*Output:* A modified version of each of the texts created and stored under the name ‹text›.del—i.e., the extension .del is added to the original text name. For the example just given, the output file would be "tailed.del."

TRANSLATE: A program for deleting concepts from a text.

*Input:* 1. A file containing a translation thesaurus. This file has the form, one concept per line with the general concept into which the following concepts are translated identified by a leading "$." A possible thesaurus file might contain:

```
$search
tailed
find
followed
$failed
```

```
elude
lose
$businessman
merchant
store-owner
$crooks
robbers
thiefs
```

2. A file containing the list of texts that the researcher wants translated.

3. A set of files, one for each text, on which the researcher wants to perform the translation. For example, one such text might be the text "tailed" previously mentioned.

*Output:* A modified version of each of the texts created and stored under the name ‹text›.tr—i.e., the extension .tr is added to the original text name. For the example given above, the output file would be "tailed.tr." It would contain the text "Killishandra and Ezra search the businessman through many corridors in order to search the crooks' lair. Unfortunately, he managed to failed them."

CMATRIX2: A program for counting the number of times concepts occur within texts.

*Input:* 1. A file containing the concepts that the researcher wants to look for in the texts. This file might contain the concepts—I, scientists, research, hard work, collaboration, discoveries, new ideas, formal, informal, co-authors, and paper.

2. A file containing the list of texts that the researcher wants to translate.

3. A set of files, one for each text, on which the researcher wants to perform a content analysis. For example, one such text might be "Student A" and another might be "Student-B." The texts for Student A and Student B would be the passages previously described.

4. Information on whether the researcher wants the program to simply denote that the concept occurs in the text (occurrence mode) or to count the number of times that the concept occurs (frequency mode).

*Output:* 1. A file *cmatrix2.dat* is created, containing a matrix such that each row represents a text (in the order in which they occur in the list of texts file) and each column a concept (in the order

in which the concepts occur in the concepts file). The entries in this matrix depend on the mode in which CMATRIX2 is used—occurrence or frequency. In the occurrence mode, each entry is a "1" if that concept occurs in that text and a "0" otherwise. In the frequency mode, each entry is the number of times that the concept occurs in that text. For the student example, cmatrix2.dat would contain the following information when the occurrence mode was chosen:

```
1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1
```

Alternatively, when the frequency mode is chosen, cmatrix2.dat would appear as:

```
1 4 2 1 2 2 2 1 1 1 1
1 4 2 1 2 2 2 1 1 1 1
```

2. The file *cstats2.dat* contains general statistics on concept usage.

### *Map Analysis*

Most of the procedures within the MECA package aid the researcher in doing map analysis. Only those procedures directly relevant to this paper will be described. These are STARTUP, CODEMAP, COMPRA, DRAWMAP, and SKI.

STARTUP: A program for defining a coding template.

*Input:* The STARTUP program is interactive, asking the researcher a set of questions about what coding choices are being made. The choices are recorded in a setup file, which then affects the operation of most other MECA programs. The researcher should have made the following decisions ahead of time: (1) the name of the file that will contain the coding choices (e.g., student.set); (2) the number of concept categories; (3) whether statement directionality will be imposed; (4) how strength will be used; and (5) the number of relationship types. It is not necessary to know apriori the number of concepts per category or the names of the concepts in those categories.

*Output:* A text file containing the coding choices. This is a text file with the name identified by the researcher (e.g., student.set).

CODEMAP: A program for assisting the researcher in coding the text.

CODEMAP is a semiautomated tool for coding maps. This tool is a new and improved version of the tool CODEF discussed in Carley (1986, 1988).

*Input:* The CODEMAP program is interactive, asking the researcher a set of questions that lead the researcher through coding the text one statement at a time. The researcher should have the following material ready: (1) the setup file (e.g., student.set); (2) the name for the file that will contain the coded map (e.g., sa.map); and (3) the text file that is being coded. The text file is not directly analyzed by CODEMAP; rather, the researcher reads the text and decides, sentence by sentence, what statements to code.

*Output:* The output is a text file with one line per coded statement. After the entire text from Student A is coded, the resultant output file would contain the following information:

```
created by researcher 11-91:
1$concept$I$concept$scientists$1
1$concept$scientists$concept$research$2
1$concept$scientists$concept$new ideas$1
1$concept$scientists$concept$co-authors$1
1$concept$research$concept$scientists$2
1$concept$collaboration$concept$scientists$1
1$concept$collaboration$concept$discoveries$1
1$concept$collaboration$concept$formal$1
1$concept$discoveries$concept$scientists$1
1$concept$research$concept$collaboration$1
1$concept$research$concept$hard work$1
1$concept$research$concept$discovery$1
1$concept$research$concept$new ideas$1
1$concept$informal$concept$new ideas$1
1$concept$formal$concept$co-authors$1
1$concept$formal$concept$paper$1
1$concept$co-authors$concept$paper$2
1$concept$paper$concept$co-authors$2
```

The first number indicates the strength, the second and fourth positions denote the concept categories, the third and fifth positions

denote the concepts, and the last number indicates the directionality. For example, line 3, the statement linking scientist to research has a "2" as the last piece of information, thus indicating that there is a bidirectional relation.

COMPRA: A program for comparing two maps.

COMPRA can be used to compare and/or combine two maps. The program generates a comparison in terms of just concepts and in terms of relationships. It does this first across all concept types and relationship types and then by each type separately. It assumes that the number of possible concepts is the number in the setup file.

*Input:* 1. Two map files that the researcher wishes to compare.

2. A setup file.

*Output:* As output COMPRA can create any or all of the following output files: (1) a map file containing the symmetric difference map for the first map; (2) a map file containing the symmetric difference map for the second map; (3) a map file containing the intersection map; (4) a map file containing the union map; (5) a map file containing statistics on the comparison of the two maps.

DRAWMAP: A program for converting a map data base into a picture. Generates a graph by placing the concepts around a circle and placing a line between each pair of concepts that occur in a statement. It can, though need not, place arrows on the lines to indicate directionality, and it can label the lines according to their strength or show strength by the boldness of the lines. Sign can be shown through either dotted lines or by labeling the lines. It ignores concept type and relationship type information.

*Input: DRAWMAP requires as input a coded map file. Optional input files include a position file that lists the location of each concept around the circle and an abbreviation file that lists alternative names for each concept.*

*Output:* The output is a file in MacDraw format.

SKI: A simple expert system for explicating implicit knowledge in a coded map. This program is described in detail in Carley (1988). This program operates entirely in batch mode.

*Input:* Four files are required as input, each in the same format as the map. These files are the map to be converted, the set of definitives, the set of logicals, and the set of simples. Definitives are statements that are presumed to embody whatever information is socially

shared. Logicals and simples contain information that though social may not have the same level of agreement.

*Output:* The output is a modified version of the map containing the new social statements.

## REFERENCES

Blank, G., J. L. McCartney, and E. Brent, eds. 1989. *New Technology and Sociology: Practical Applications in Research and Work.* New Brunswick, N.J.: Transaction Publishers.

Bobrow, D. G., and A. Collins, eds. 1976. *Representation and Understanding: Studies in Cognitive Science.* New York: Academic Press.

Brachman, R. J., and H. J. Levesque, eds. 1985. *Readings in Knowledge Representation.* Los Altos, Calif.: M. Kaufmann.

Bruce, B., and D. Newman. 1978. *Interacting Plans* (Tech. Rep.). Champaign, Ill.: Center for the Study of Reading.

Carley, K. M. 1984. "Constructing Consensus." Ph.D. diss., Harvard University.

———. 1986. "An Approach for Relating Social Structure to Cognitive Structure." *Journal of Mathematical Sociology* 12(2): 137–89.

———. 1988. "Formalizing the Social Expert's Knowledge." *Sociological Methods and Research* 17:165–232.

———. 1990. "Computer Analysis of Qualitative Data—Copy of Overheads for Didactic Seminar." Paper presented at annual meeting of the American Sociological Association.

———. Forthcoming *b.* "The Social Construction of Knowledge." In *Experiments in Epistemology,* edited by R. Lawler. Norwood, N.J.: Ablex.

———. Forthcoming *a.* "Content Analysis." In *The Encyclopedia of Language and Linguistics,* edited by R. E. Asher et al. Edinburgh: Pergamon Press.

Carley, K., and D. Kaufer. Forthcoming. "Semantic Connectivity: An Approach for Analyzing Semantic Networks." *Communication Theory.*

Carley, K., and M. Palmquist. 1992. "Extracting, Representing, and Analyzing Mental Models." *Social Forces* 70:(3) 601–36.

Cicourel, A. V. 1974. *Cognitive Sociology.* New York: Free Press, Macmillan.

Cicourel, A. V., and K. Carley. 1990. "The Coder of Narrative as Expert: The Ecological Validity of Coding Practices." Paper presented at the annual meeting of the American Anthropological Association.

Conrad, P., and S. Reinharz. 1984. "Computers and Qualitative Data." *Qualitative Sociology* 7:3–15.

Danowski, J. A. 1980. "Computer-Mediated Communication: A Network-based Content Analysis Using a CBBS Conference." In *Communication Yearbook,* edited by M. Burgoon, 905–24. Beverly Hills: Sage.

———. 1982. "A Network-based Content Analysis Methodology for Computer-Mediated Communication: An Illustration with a Computer Bulletin Board." In *Communication Yearbook,* edited by R. Bostrom, 904–25. New Brunswick, N.J.: Transaction Books.

———. 1988. "Organizational Infographics and Automated Auditing: Using Computers to Unobtrusively Gather and Analyze Communication." In *Handbook of Organizational Communication,* edited by G. Goldhaber and G. Barnett, 385–433. Norwood, N.J.: Ablex.

Dietrich, R., and C. F. Graumann. 1989. *Language Processing in Social Context.* New York: Elsevier Science Publishing.

Dunphy, D.C., C.G. Bullard, and E.E.M. Crossing. 1974. "Validation of the General Inquirer Harvard IV Dictionary." Paper presented at 1974 Conference on Content Analysis, Pisa, Italy.

Dyer, M. 1983. "The Role of Affect in Narratives." *Cognitive Science* 7:211–42.

Ellis, F. S. 1968. *A Lexical Concordance to the Poetical Works of Percy Bysshe Shelley.* New York: B. Franklin.

Ericsson, K. A., and H. A. Simon. 1984. *Protocol Analysis: Verbal Reports as Data.* Cambridge: MIT Press.

Fan, D. 1988. *Predictions of Public Opinion From the Mass Media: Computer Content Analysis and Mathematical Modeling.* New York: Greenwood Press.

Fauconnier, G. 1985. *Mental Spaces: Aspects of Meaning Construction in Natural Language.* Cambridge: Bradford Books, MIT Press.

Franzosi, R. 1990*a*. "From Words to Numbers: A Generalized and Linguistic-based Coding Procedure for Collecting Textual Data." In *Sociological Methodology,* edited by C. Clogg. San Francisco: Jossey-Bass.

———. 1990*b*. "Computer-Assisted Coding of Textual Data: An Application to Semantic Grammars." *Sociological Methods and Research,* 19:(2)225–57.

———. 1990*c*. "Strategies for the Prevention, Detection, and Correction of Measurement Error in Data Collected from Textual Sources." *Sociological Methods and Research* 18:(4)442–72.

Gallhofer, I. N., and W. E. Saris. 1988. *Sociometric Research VI: Data Collection and Scaling.* New York: St. Martin's Press.

Garson, D. 1985. "Content analysis machine-readable data file V 1.0." Raleigh, N.C.: National Collegiate Software Clearinghouse.

Gentner, D., and A. L. Stevens, eds. 1983. *Mental Models.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Grey, A., D. Kaplan, and H. D. Lasswell. 1965. "Recording and Context Units–Four Ways of Coding Editorial Content." In *Language of Politics,* edited by H. D. Laswell and N. Leites, 113–26. Cambridge: MIT Press.

Holsti, O. R. 1954. Content Analysis. In *The Handbook of Social Psychology,* vol. II, edited by G. Lindzey and E. Aronson, 596–692. Reading, Mass.: Addison-Wesley.

———. 1969. *Content Analysis for the Social Sciences and Humanities.* Reading, Mass.: Addison-Wesley.

Hutchins, W. J. 1982. "The Evolution of Machine Translation Systems." In *Practical Experience of Machine Translation: Proceedings of a Conference, London Nov. 1981,* edited by V. Lawson, 21–38. New York: North-Holland.

Johnson-Laird, P. 1983. *Mental Models: Toward a Cognitive Science of Language, Inference, and Consciousness.* Cambridge: Harvard University Press.

Kaufer, D. S., and K. M. Carley. 1993. *Communication at a Distance: The*

*Influence of Print on Sociocultural Organization and Change*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Krippendorf, K. 1980. *Content Analysis: An Introduction to its Methodology*. Beverly Hills: Sage.

Lehnert, W. 1981. "Plot Units and Narrative Summarization." *Cognitive Science* 5:293–331.

Lehnert, W., and E. Vine. 1987. "The Role of Affect in Narrative Structure." *Cognition and Emotion* 1:299–322.

Lindkvist, K. 1981. "Approaches to Text Analysis." In *Advances in Content Analysis*, edited by K. E. Rosengren, 23–42. Beverly Hills: Sage.

Mallery, J. 1985. "Universality and Individuality: The Interaction of Noun Phrase Determiners in Copular Clauses." *Twenty-third Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference*. Chicago.

Mallery, J., and G. Duffy. 1986. *A Computational Model of Semantic Perception* (Tech. Rep. AI memo no. 799). Cambridge: MIT Artificial Intelligence Laboratory.

Mandler, J. M. 1984. *Stories, Scripts, and Scenes: Aspects of Schema Theory*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

McTavish, D. G., and E. B. Pirro. 1990. "Contextual Content Analysis." *Quality and Quantity* 24:245–65.

Merleau-Ponty, M. 1964. *Consciousness and the Acquisition of Language*. Evanston, Ill.: Northwestern University Press.

Namenwirth, J., and R. Weber. 1987. *Dynamics of Culture*. Boston: Allen & Unwin.

Neuman, W.R. 1989. "Parallel Content Analysis: Old Paradigms and New Proposals." *Public Communication and Behavior* 2:205–89.

Ogilvie, D.M., P.J. Stone, and E.F. Kelly. 1982. Computer-Aided Content Analysis." In *A Handbook of Social Science Methods*, vol. 2, edited by R. B. Smith and P. K. Manning, 219–46. New York: Ballinger.

Osgood, C. E. 1959. "The Representational Model and Relevant Research Methods." In *Trends in Content Analysis*, edited by I. S. Pool, 33–88. Urbana: University of Illinois Press.

Palmquist, M. E. 1990. "The Lexicon of the Classroom: Language and Learning in Writing Classrooms." Ph.D. diss. Carnegie Mellon University, Pittsburgh, PA.

Polanyi, L. 1985. *Telling the American Story: A Structural and Cultural Analysis of Conversational Storytelling*. Norwood N.J.: Ablex.

Polanyi, M. P. 1962[1958]. *Personal Knowledge: Towards a Post-Critical Philosophy*. Chicago: University of Chicago Press.

Potter, R. G. 1982. "Reader Responses and Character Syntax." In *Computing in the Humanities*, edited by R. W. Bailey, 65–78. New York: North-Holland.

Qualis Research Associates. 1989. *The Ethnograph*. Littleton, Colo.:

Roberts, C. W. 1989. "Other Than Counting Words: A Linguistic Approach to Content Analysis." *Social Forces* 68:147–77.

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) 313–38.

Saburi-Haghighi, N. 1991. "World Bank's Adjustment Lending Policy 1980–1989: Towards Understanding the Socio-Political Dimensions of Adjustment." Ph.D. diss. American University, Washington, D.C.

Sacks, H. 1972. "On the Analysability of Stories by Children." In *Directions in Socio-linguistics,* edited by J. Gumprez and D. Hymes, 329–45. New York: Holt, Rinehart & Winston.

Saris-Gallhofer, I. N., W. E. Saris, and E. L. Morton. 1978. "A Validation Study of Holsti's Content Analysis Procedure." *Quality and Quantity* 12:131–45.

Schank, R. C., and R. Abelson. 1977. *Scripts Plans and Goals and Understanding.* New York: Wiley.

Schank, R. C., and C. K. Riesbeck, eds. 1981. *Inside Computer Understanding.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Sowa, J. F. 1984. *Conceptual Structures.* Reading, Mass.: Addison-Wesley.

Sozer, E. 1985. *Text Connexity, Text Coherence: Aspects, Methods, Results.* Hamburg, Germany: Buske.

Splichal, S., and A. Ferligoj. 1988. "Ideology in International Propaganda." In *Sociometric Research V1: Data Collection and Scaling,* edited by W. E. Saris and I. N. Gallhofer, ch. 5. New York: St. Martin's Press.

Sproull, L. S., and R. F. Sproull. 1982. "Managing and Analyzing Behavioral Records: Explorations in Non-Numeric Data Analysis." *Human Organization* 41:283–90.

Stone, P. J., D. C. Dunphy, M. S. Smith, and D. M. Ogilvie. 1968. *The General Inquirer: A Computer Approach to Content Analysis.* Cambridge: MIT Press.

Stone, P. J., and Cambridge Computer Associates. 1968. *User's Manual for the General Inquirer.* Cambridge: MIT Press.

Stubbs, M. 1983. *Discourse Analysis: The Sociolinguistic Analysis of Natural Language.* Chicago: University of Chicago Press.

Sullivan, M. P. 1973. *Perceptions of Symbols in Foreign Policy: Data from the Vietnam Case* (Tech. Rep.). Ann Arbor: Inter-University Consortium for Political Research.

Thomas, D. 1991. "Changing Cultural Perception of Robots As Seen in Science Fiction." Bachelor's thesis, Carnegie Mellon University.

VanLehn, K., and S. Garlick. 1987. "Cirrus: An Automated Protocol Analysis Tool." In *Proceedings of the Fourth International Workshop on Machine Learning,* edited by P. Langley, 205–17. Los Altos, Calif.: Morgan-Kaufman.

Van Meter, K. M., and L. Mounier. 1989. "East Meets West: Official Biographies of Members of the Central Committee of the Communist Party of the Soviet Union Between 1981 and 1987, Analyzed with Western Social Network Analysis Methods." *Connections* 12:32–38.

Weber, R. 1984. "Computer-Aided Content Analysis: A Short Primer." *Qualitative Sociology* 7:126–47.

Woodrum, E. 1984. " 'Mainstreaming' Content Analysis in Social Science: Meth-

odological Advantages, Obstacles, and Solutions." *Social Science Research* 13:1–19.

Young, R., W. B. Stevenson, T. Nicol, and G. W. Gilmore. 1936. *Analytical Concordance to the Bible on an Entirely New Plan Containing About 311,000 References, Subdivided under the Hebrew and Greek Originals, with the Literal Meaning and Pronunciation of Each Designed for the Simplest Reader of the English Bible.* New York: Funk and Wagnalls.