Request ID:

Incoming Request

ID: 5669102
STATUS: New   SERVICE TYPE: Copy
EXPIRY DATE: 6/10/2005
NOTE: ARIEL 128.2.114.135 **New as of 8/26/04**
FAX/ARIEL:(412)681-1998 EMAIL:es82@andrew.cmu.edu
COPYRIGHT: CCG
REQUESTER: (OCLC:PMC) (illhost.oclc.org:1611)
SUPPLIER: (OCLC:ZYU) (illmgr.bobst.nyu.edu:1611)
REQUEST DATE: 6/1/2005 2:57:45 PM
NEED BEFORE DATE: 6/29/2005
MAXIMUM COST: 35.00 (USD)
PAYMENT TYPE: IFM
ITEM TYPE/MEDIUM: Serial
REQUESTER NUMBER: (OCLC) ILLNUM:5669102
ISO TRANSACTION: OCLC:PMC-5669102:ZYU

TITLE: Text analysis for the social sciences : methods for drawing statistical inferences from texts and transcripts /
PUBLISHED: Mahwah NJ. : Erlbaum, c1997. :
AUTHOR OF ARTICLE: Michael Palmquist, Kathleen M. Carley, Thomas Dale: Two Applications of Automated Text Analysis: Analyzing Literary and Non-Literary Texts
VOLUME/ISSUE/DATE: (1997)
PAGES: 171-189
SERIES: LEA's communication series
BIB RECORD ID: OCLC:36103959
SOURCE: <TN:117271> OCLC[ISBN:0805817344 (alk. pap]

PATRON NAME: Economou, Rochelle

DELIVERY DETAILS

REQUESTED DELIVERY MODE: IDS #187A01/Library Mail
PHYSICAL DELIVERY ADDRESS:
ILL/E&S Library/Wean Hall/Carnegie Mellon University/5000 Forbes Ave/Pittsburgh/PA/15213-3890

BILLING DETAILS

BILLING ADDRESS:
Same **FEIN 25-0969-449**

TEXT ANALYSIS FOR THE SOCIAL SCIENCES

Methods for Drawing Statistical Inferences From Texts and Transcripts

Edited by

CARL W. ROBERTS
Iowa State University

LEA
1997
LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
Mahwah, New Jersey

Lawrence Erlbaum Associates, Inc., Publishers
10 Industrial Avenue
Mahwah, New Jersey 07430

Library of Congress Cataloging-in-Publication Data

Text analysis for the social sciences : methods for drawing statistical inferences from texts and transcripts / edited by Carl W. Roberts.
p. cm.
Includes bibliographical references and indexes.
ISBN 0-8058-1734-4 (alk. paper). — ISBN 0-8058-1735-2 (pbk. alk. paper).
1. Social sciences—Methodology. 2. Content analysis (Communication). 3. Social sciences—Research—Methodology.
I. Roberts, Carl W.
H61.T456 1997
300'.1—dc21

96-40299
CIP

Books published by Lawrence Erlbaum Associates are printed on acid-free paper, and their bindings are chosen for strength and durability.

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

CONTENTS

CHAPTER

10

APPLICATIONS OF COMPUTER-AIDED TEXT ANALYSIS: ANALYZING LITERARY AND NONLITERARY TEXTS

Michael E. Palmquist
Colorado State University

Kathleen M. Carley
Thomas A. Dale
Carnegie Mellon University

This chapter describes a methodology for representing mental models as maps, extracting those maps from texts, and analyzing and comparing the maps. Drawing on a computer-based cognitive mapping procedure developed by Carley (1986a, 1988; Carley & Palmquist, 1992), we illustrate the methodology with two analyses—one of 27 works of science fiction in which robots play a significant role, the other of students' understanding of class material during a course on academic writing. Results indicate that by the 1980s robots were portrayed much more positively than they were prior to the 1940s. Evidence is also provided that during the semester students gained consensus in their understanding of academic writing. The chapter concludes with a discussion of potential applications of the methodology.

In this chapter, we describe and discuss two studies that employ computer-aided text analysis. The first study explores depictions of robots in 27 works of science fiction published between 1818 and 1988 (Dale, 1991). The second study explores the growth of shared knowledge about writing among students in an introductory college composition classroom (Carley & Palmquist, 1992; Palmquist, 1990). The two studies share a number of similarities. Both employ computer-aided text analysis. Both develop their concept lists, or dictionaries, empirically (i.e., based on concepts actually found in the texts under analysis). Both employ independent raters to assess the reliability of human judgments concerning coding of concepts and statements (or pairs of related concepts). Both employ statistical tests to explore trends in the data.

Despite these similarities, the two studies differ in important ways. The robot study is an analysis of literary texts (specifically, plays, short stories, and novels), whereas the writing study is an analysis of student interviews and written journal entries. In addition, the writing study includes a comparison of observable behaviors (i.e., student interaction) with data drawn from texts, whereas the robot study deals almost entirely with data drawn from texts. Finally, the two studies differ in that only one specifies how concepts are related in the texts. The robot study defines six types of relations between a single central concept *robot* and more than 400 other concepts in a concept list. In contrast, over 200 concepts are considered in the writing study, but the types of relations among those concepts are left undefined. The similarities that extend across the two studies allow us to generalize about how computer-aided text analysis might be applied in other research contexts. In turn, the differences provide us with examples of the different uses to which computer-aided text analysis might be put.

STUDY ONE: DEPICTIONS OF ROBOTS IN SCIENCE FICTION

This study analyzed 27 works of science fiction by 20 authors (for a more complete discussion of this study, see Dale, 1991). In each of the selected plays, short stories, and novels, robots play a significant role. The texts ranged from Shelly's *Frankenstein* and Capek's *R.U.R.* to Asimov's *The Robots of Dawn* and Anthony's *Robot Adept*, with particular attention paid to texts published during the past five decades. In all, 30 robots are depicted in the texts.

The texts were selected using a three-step process. First, an electronic database search of the holdings of Hunt Library at Carnegie Mellon University was conducted using the terms *Science Fiction* and *Robot$*, with the *$* denoting a wild-card character. Second, the results of the search were screened to produce a preliminary set of 65 texts that matched the search criteria. Finally, the researcher selected a subset of 27 texts on the basis of their representativeness and, to a lesser extent, on the relative importance of the author. Mary Shelly's *Frankenstein*, for instance, was selected because it is one of the first literary works that explores the implications of creating an artificial being. Several of Isaac Asimov's works were selected because he is both a seminal science fiction writer and a writer well known for using robots as central characters in his stories. Because no attempt was made to create a random sample of works of science fiction in which robots played significant roles in the plot, we present the results of this study strictly as a demonstration of the ways in which such texts can be analyzed, rather than as generalizable findings.

Our analysis of these texts was designed to explore whether we could measure differences in attitudes exhibited toward robots by authors writing at different times. To pursue that goal, six questions were addressed. Do the types of robots that authors describe (e.g., android, metallic humanoid) change over time? Do the features of robots change over time? Do the actions in which robots engage change over time? Do the emotions that robots exhibit change over time? Do the attributions that characters in the play, short story, or novel make to the robots change over time? Does the overall depiction of the robots in each work of fiction (i.e., positive or negative) change over time?

Method

Grouping the Texts. To determine whether attitudes toward robots changed over time, the texts were placed into three groups. The two dividing points were chosen in light of two important milestones in the U.S. space program. The first, 1960, marks the beginning of the decade in which the United States, spurred first by the Soviet success of Sputnik and then by John Kennedy's challenge to send a mission to the moon by the end of the decade, firmly committed itself to the space race. The second point, 1969, marks the successful lunar landing of Apollo 11. The 1960s is seen as a pivotal decade not only for the development of the U.S. space program but also for the national self-image of the United States as a high technology culture. Using 1960 and 1969 as demarcation points, the 30 maps of robots can be grouped into one group of 12 robots in texts written before 1960, one group of 6 robots in texts written between 1960 and 1969, and one group of 12 robots in texts written from 1970 to 1988.

Creating the Concept List and Relationship Types. Cognitive maps are networks of statements, where each statement consists of a pair of related concepts. Thus, prior to encoding statements, one must determine the concepts and relations according to which texts are to be encoded. The concept list and relationship types for this study were developed empirically; that is, they were established during and after the texts were read rather than prior to reading the texts. (For a more complete discussion of issues related to creating concept lists, see Carley & Palmquist, 1992.) Following an initial reading of the texts, six types of relationships were established: (1) the type of robot in each text (e.g., "Robot <IS OF THIS TYPE> metallic nonhumanoid"); (2) the features each robot possessed (e.g., "Robot <HAS FEATURE> eyes"); (3) the actions each robot engaged in (e.g., "Robot <DOES ACTION> walking"); (4) the emotions each robot experienced (e.g., "Robot <HAS EMOTION> anger"); (5) attributions made to each robot by characters in each text (e.g., "Robot <HAS CHARACTERISTIC ATTRIBUTED BY CHARACTER> pride"); and (6) the author's apparent attitude toward each robot, ranging from very negative to very positive (e.g., "Robot <IS EVALUATED BY AUTHOR> positive"). In all, the researcher coded 412 concepts as being related in one of these six ways to the robots in the set of texts.

Creating Maps. To answer each of the research questions, the texts were coded and subsequently analyzed using MECA, or Map Extraction Comparison and Analysis.[1] One researcher coded the entire set of texts, producing maps of 30 robots (three texts contained two major characters who were robots; the others contained one). A second researcher coded a subset of three texts. Using signal-detection analysis (Lindsay & Norman, 1972, pp. 664–682) agreement between the two raters was 97.5% for concepts and 97.0% for statements.[2] False alarms (defined as instances in which the second rater coded a concept or statement as being present when the first rater had not coded it as being present) were 15.1% for concepts and 40.4% for statements. The high false alarm rate for statements reflects the tendency of the second rater to create more detailed maps of each text. For the three texts, the total number of statements coded by the second rater exceeded that of the first rater by roughly 25% to 40%.

After the coding was completed, a separate set of four coders ranked the emotions and character attributions identified by the primary coder. The various emotions and attributions were ranked as either negative, neutral, or positive. These rankings afforded the three-level ordinal measure on these variables used in subsequent statistical analyses. Average simple agreement among the raters was 80.8%.

Creating Modified Maps Using SKI. Because the goal of this study was to represent the depictions of each robot as accurately as possible, a software program, SKI, was used to make implicit statements explicit. SKI accepts two files as input, a map file and what might be termed an implicit knowledge file. A fairly simple illustration may clarify the manner in which SKI operates. Imagine that the implicit knowledge file contains a statement amounting to "if something has ears then it can hear." Then, if a map file contained the statement that a robot has ears, SKI would add two additional statements: if a robot has ears then a robot can hear, and a robot can hear. When creating modified maps, SKI adds concepts and statements that explicate the implicit knowledge. SKI does not delete concepts. It can add concepts and add relations among existing or new concepts, thus creating statements.

---

[1]MECA is a collection of 15 text analysis programs. The programs can run on UNIX workstations, many IBM personal computers or clones, and Apple Macintosh computers (Carley, 1986a, 1988, 1993; Carley & Palmquist, 1992). For availability, see Popping (this volume).

[2]To increase comparability of statements across texts, a number of explicit concepts were coded as synonyms. For instance, the explicit concepts "annihilated" and "destroyed," as in "the robot annihilated the humans," were coded as "attack."

Although SKI cannot overcome gross coder error, it can overcome some of the most frequent coder errors, such as neglecting to fully specify all implied statements in the text. The modified maps generated by SKI are thus larger than those generated by novice coders. SKI-modified maps tend to be closer to maps coded by individuals who are experts in the domain from which the texts are selected. Carley (1988) described SKI in detail and demonstrated how its application to texts coded by novices actually improves the coding so that it more closely resembles coding by experts.

A comparison of agreement on features of robots between unmodified and SKI-modified maps indicates that the modified maps allow more common features to be identified (see Table 10.1). Similar results were obtained for emotions, actions, and character attributions. Because of this, the subsequent analyses are based on maps modified using SKI.

Analyzing Maps. After each map had been modified by SKI, all maps were analyzed statistically and depicted graphically. To prepare the maps for statistical analysis, software was used to produce data matrices for each map in which the rows were either concepts or statements and the columns were the names of the other maps. The cells in the data matrices indicated whether or not each concept or statement was found in that map. Summary matrices

TABLE 10.1
Robot Features by Period

	Pre-1960s (12 texts)		*1960s (6 texts)*		*Post-1960s (12 texts)*	
Unmodified maps						
	Hear	(12)	Hear	(4)	Hear	(10)
	Memory	(9)	Memory	(3)	Memory	(12)
	Voice	(10)	Voice	(4)	Voice	(10)
			Circuits	(5)	Circuits	(10)
			Consciousness	(4)	Consciousness	(10)
SKI-modified maps						
	Brain	(12)	Brain	(6)	Brain	(12)
	Face	(10)	Face	(4)	Face	(10)
	Hear	(12)	Hear	(4)	Hear	(12)
	Intelligence	(12)	Intelligence	(6)	Intelligence	(12)
	Memory	(9)	Memory	(3)	Memory	(12)
	Senses	(12)	Senses	(4)	Senses	(12)
	Sight	(9)	Sight	(3)	Sight	(9)
	Voice	(11)	Voice	(4)	Voice	(10)
	Eyes	(9)			Eyes	(9)
			Circuits	(5)	Circuits	(10)
			Consciousness	(4)	Consciousness	(10)

Note. The numbers in parentheses indicate the number of maps in which each feature is present.

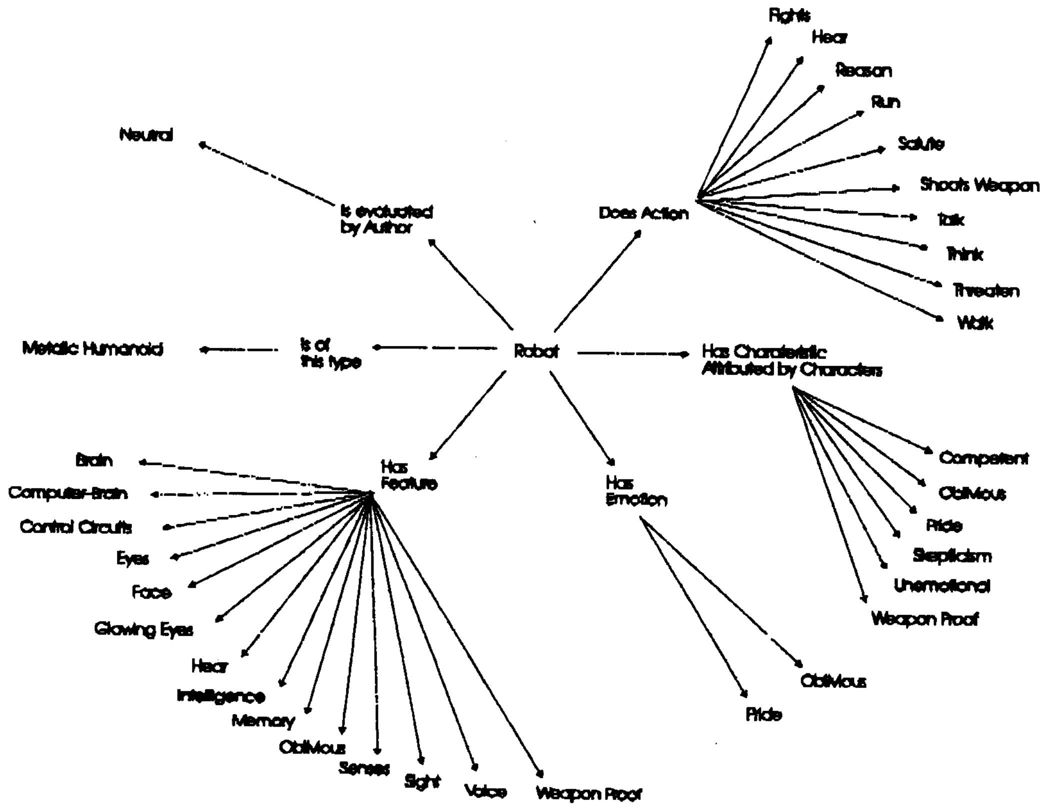

FIG. 10.1. Map of "Arm of the Law."

were subsequently produced in which the rows were individual maps and the columns were variables indicating the total numbers of concepts and statements in each map, the intersection of concepts and statements across the set of maps, the number of concepts and statements in each map not found in each of the other maps, and the number of concepts and statements that were unique (i.e., not found in any of the other maps) to each map.

Maps were also depicted graphically. A modified map of Harry Harrison's short story, "Arm of the Law," is presented in Figure 10.1. To read the map as a series of statements, begin at the central concept, "robot," read a phrase such as DOES ACTION as the relationship between robot and another concept, and read one of the concepts linked to DOES ACTION by a line as the second concept. The arrow on each line indicates the direction in which the statement should be read. For example, you could read "Robot <DOES ACTION> Fights" as a statement meaning that the robot in "Arm of the Law" fought. In a more detailed map, concepts such as "fight" might be linked to still other concepts, such as the person or thing being fought, the type of weapons (if any) used in the fight, and who won the fight. The reader should note as well that a map of a longer text, such as Robert Heinlein's novel, *The Moon is a Harsh Mistress*, would be more detailed than the map of "Arm of the Law."

Results

Comparisons of Figures 10.2, 10.3, and 10.4 indicate only one trend across the three time periods in robot types, features, or actions. Over time, fewer authors depicted robots as nonmetallic humanoids, and more authors depicted robots as metallic humanoids. This trend is mirrored by the presence of one feature, circuits, in the second and third periods that does not appear in the first period.

Clearer trends over time can be seen in the emotions that robots exhibit during the three periods, the attributions made to robots by characters in the texts, and the overall evaluations of robots by the authors of the texts. In general, the characterizations of robots in texts written after the 1960s are more positive than those written prior to or during the 1960s. In the texts written prior to 1960, consensus exists only on the emotions "disdain" and "fear" and on the character attributions "disdain," "fear," "formidable," and "intelligent." In the 1960s, the emotion "disdain" is joined by "anger," whereas "fear" is not present, and the character attribution "disdain" is joined by "anger" and "pride," but "fear" is not present. In the 1970s and 1980s, "fear" is present but is counterbalanced by the more positive emotions of "loyalty," "trust," and "friendship," emotions that authors do not attribute to robots in the first two periods. Similarly, in the 1970s and 1980s, the character attribution "fear" is present but is counterbalanced by the

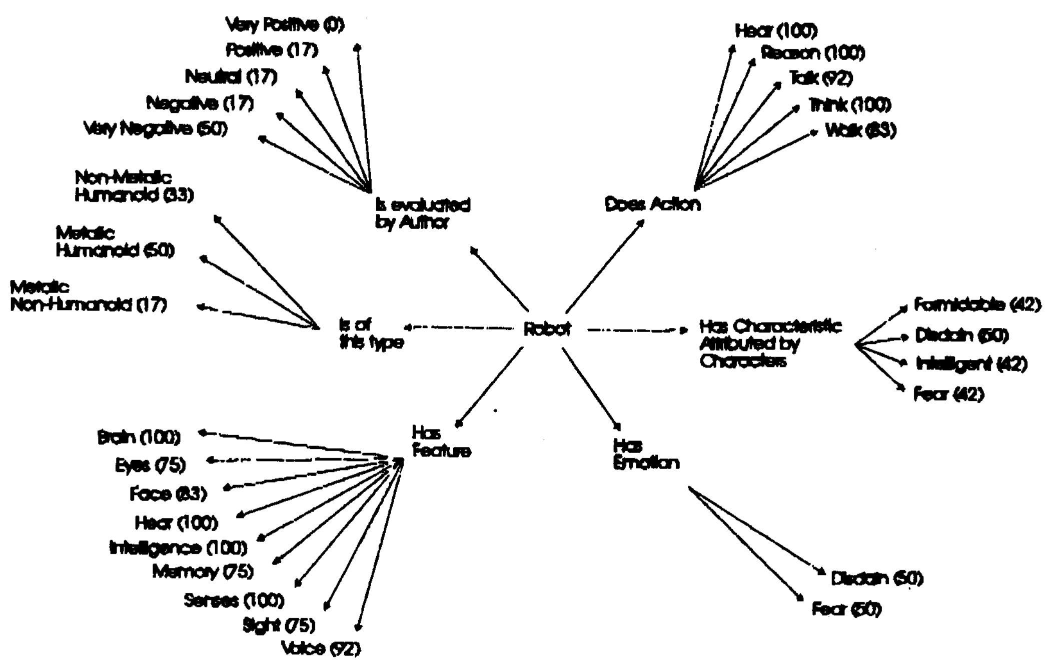

FIG. 10.2. Map of pre-1960s texts. In parentheses are percentages of texts in which statement occurs.

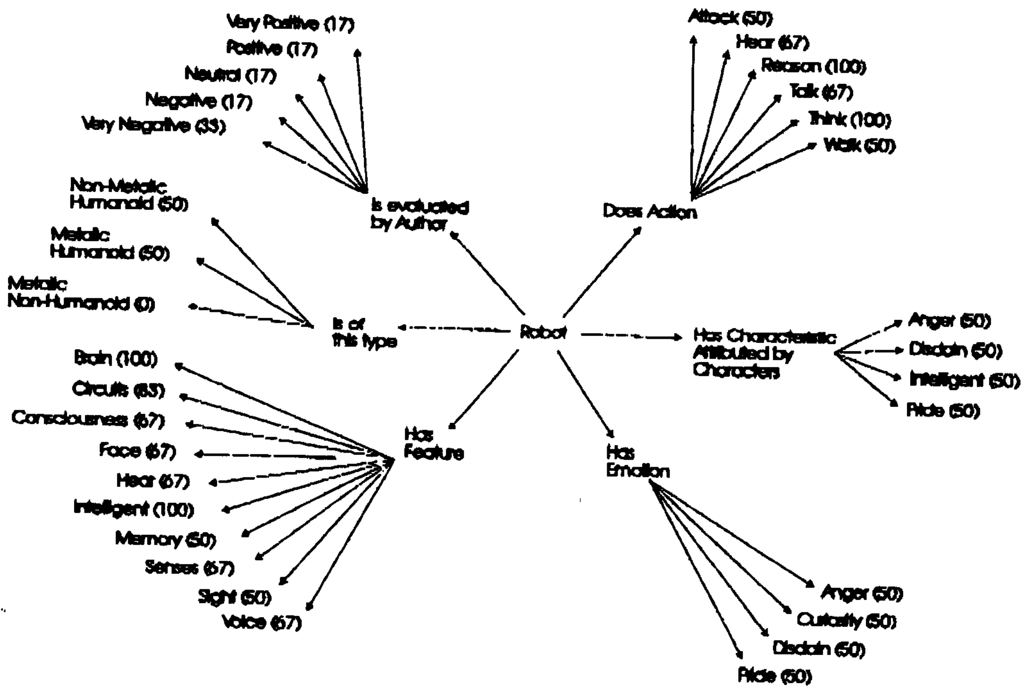

FIG. 10.3. Map of 1960s texts. In parentheses are percentages of texts in which statement occurs.

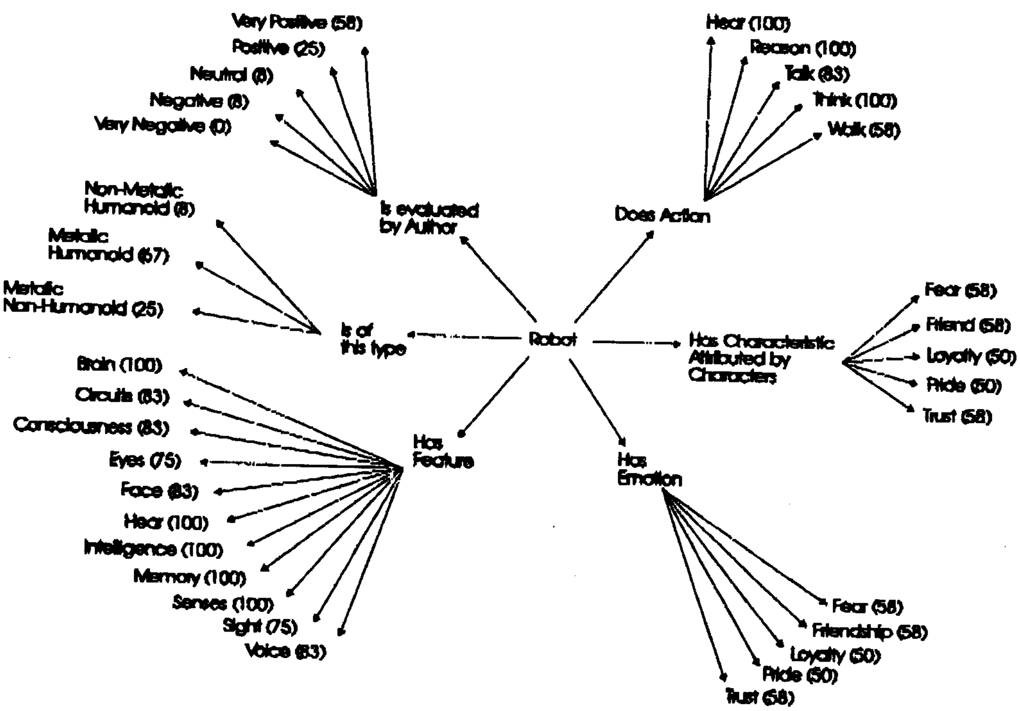

FIG. 10.4. Map of post-1960s texts. In parentheses are percentages of texts in which statement occurs.

more po
ties that
Statis
emotion,
for the
rankings
variance
in the m
were exl
characte
scores f
evaluati
than for
A Pea
ments a
1818 and
trend (r
of positi
between
publicat
positive
percent

Discuss

Resul
robots b
robots h
emotion

Robot em
Character
Author ev

Note.
are mean
deviation
*p < .(

more positive character attributions of "loyalty," "trust," and "friend," qualities that characters do not attribute to robots in the first two periods.

Statistical analysis mirrors these trends. Positive and negative robot emotion, character attribution, and author evaluation scores were calculated for the modified map of each text, the first two of which were based on rankings by four raters. Table 10.2 presents the results of three analyses of variance that showed significant differences among the three time periods in the mean percentage of science fiction texts in which positive emotions were exhibited by robots and in which positive attributions were made by characters. The table also lists the mean author evaluation score. The mean scores for positive emotion statements, character attributes, and author evaluations were significantly higher for texts written in the 1970s and 1980s than for the other two periods.

A Pearson correlation between the percentage of positive emotion statements and publication year (excluding *Frankenstein*, which was published in 1818 and which served as a strong outlier) also indicated a significant positive trend ($r = .468$, $p < .01$). As publication year grew more recent, the percentage of positive emotion statements increased. Similarly, a Pearson correlation between the percentage of positive character attribution statements and publication year (again excluding *Frankenstein*) also indicated a significant positive trend ($r = .380$, $p < .05$). As publication year grew more recent, the percentage of positive character attribution statements increased.

Discussion

Results. These results illustrate change over time in the depictions of robots by the authors in this set of texts. Although the actions and features of robots have remained relatively constant during the periods studied, robots' emotions, characters' attributions, and authors' evaluations have shifted

TABLE 10.2
Temporal Changes in Statements Indicating Robot Emotions, Character Attributions, and Author Evaluations That Are Positive

	Pre-1960s (12 texts)	*1960s (6 texts)*	*Post-1960s (12 texts)*	*F*
Robot emotions	28.1	29.5	55.0	6.25**
	(20.9)	(24.1)	(17.8)	
Character attributions	26.4	27.5	47.1	3.58*
	(18.7)	(20.7)	(21.9)	
Author evaluations	2.0	2.7	4.3	11.43***
	(1.2)	(1.6)	(1.0)	

Note. Robot emotions and character attributions are in percentages; author evaluations are means on a 5-point scale with 1 being very negative and 5 being very positive. Standard deviations are in parentheses.

*$p < .05$. **$p < .01$. ***$p < .001$.

from largely negative to largely positive. By the 1970s and 1980s, robots were considered capable of inspiring loyalty, trust, and even friendship within their human counterparts.

Because the texts were not selected randomly, however, we cannot generalize these findings to depictions of robots in science fiction in general. Despite this limitation, these results suggest that a more ambitious investigation of texts in which robots play a significant role might reveal changes over time in writers' presentation of robots and, by implication, of changes over time on a societal level.

Method. This study provides a relatively straightforward example of how computer-aided text analysis can be used to conduct map analysis. However, at least two issues related to the use of the methodology are worth noting: the inherent threats to reliability and validity entailed by the use of human coders and the data reduction problem. First, the use of human raters to code the texts (as opposed to machine coding, which carries with it another set of challenges) raises questions about the reliability and validity of that coding. Although the interrater reliability reported in the study is at an acceptable level (i.e., over 80% agreement), analysis of the results indicates some anomalies. For instance, based on the 1960s texts our analysis of the maps indicates that three of the robots have vision, yet only two robots are listed as having eyes. The explanation for this apparent discrepancy lies in the manner in which the robot in Heinlein's *The Moon is a Harsh Mistress* "sees." Essentially, the robot is a mainframe computer that sees through video cameras. We might question whether those cameras should have been listed as eyes, and we might ask what difference exists between the video cameras and an analogous light sensor in a metallic humanoid robot. Because a human coder could reasonably choose one of several answers in this situation, this is a case in which human coding can lead to a great deal of ambiguity.

The use of human raters also introduces problems related to inconsistency in the level of analysis as a coder moves from one text to another. Particularly when dealing with a large set of texts, it is possible to have judgments being made in a qualitatively different manner at the beginning and end of coding a set of texts. This issue can be dealt with to some extent by using multiple coders and by making more than one pass through each text. However, it is difficult, when using human coders, to ensure that texts are coded in precisely the same manner.

The nature of text analysis also necessarily entails some data reduction. The amount of information provided by a statement that a robot possesses a certain feature or exhibits a certain emotion cannot be expected to provide the same amount of information conveyed by an author's careful development of a character or a scene. For instance, a statement that is coded

"Robot <DOES ACTION> Attack" does not indicate who was attacked, why they were attacked, or how they were attacked. Map analysis would be able to indicate that the same statement occurred in two or more texts, but it would not be able to indicate, unless it were explicitly coded as such, that in one case the attack occurred because the robot was defending someone whereas in another case the robot was running amok. The data reduction problem can, to some degree, be alleviated by careful and comprehensive coding of a given text. However, it seems unlikely that researchers will always want to create highly detailed maps of a particular text. For the majority of researchers, it seems that at least some data reduction is inevitable.

STUDY TWO: LANGUAGE AND LEARNING IN A UNIVERSITY WRITING CLASSROOM

A common problem in writing instruction is lack of a shared set of terms for discussing writing. Students are often uncertain, for instance, about the meaning of terms such as *voice* or *tone*, and writing instructors often express dismay about students' misunderstanding of important terms such as *revision*. When teachers suggest that a student revise a draft, for instance, they typically want their students to rethink global issues such as audience, purpose, or structure. Students, on the other hand, often interpret a suggestion to revise a paper as a need to check spelling, punctuation, and word choice. This kind of mismatch between intention and understanding can significantly reduce the learning that takes place in a classroom.

This study explores the evolution of a shared lexicon about writing among the 16 students and teacher in one semester-long writing course (for a more complete description of the study, see Palmquist, 1990). The purpose of the course was to introduce students to academic writing. Specifically, students were taught that academic writing involved reading, interpreting, and responding to arguments. One expectation of this study was that, as the semester progressed, students and teachers would develop a shared understanding of key concepts and the relations between pairs of these concepts (i.e., within statements). To determine whether that expectation was met, maps of students' mental models of writing at the beginning and end of the semester were created and analyzed. These maps were compared both with each other and with a map of the instructor's mental model of the course content. Four questions were addressed. As the semester progressed, did student maps become more similar to those of their teacher? As the semester progressed, did student maps become more similar to those of their classmates? Which concepts and statements were used by at least half the students at the beginning and end of the semester? Was student interaction positively associated with similarities among student maps?

Data and Method

Data Collection. Students were interviewed at the beginning, middle, and end of the semester. They also submitted writing journals in which they discussed key concepts in the course, among them the purpose of research writing. The teacher was interviewed at the beginning of the semester. Class sessions were observed, and key concepts used during discussions and lectures were noted.

In-class interaction was tracked through classroom observation notes and records of computer-based interactions among students during a once-weekly meeting in a computer classroom. Interactions were summarized using a grid for each week of the semester. The weekly grids were subsequently summarized into a single grid, each cell of which contained the total number of in-class interactions between each pair of students during the semester.

To track out-of-class interaction, students were asked to indicate (on a form administered each week) classmates with whom during the past week they had discussed course-related matters outside the classroom. Each interaction sheet consisted of a space for the student to write his or her name, a list of all students in the class, and a line following each name in the list. Students were asked to indicate the number of meetings they had with each classmate and to note briefly what they discussed during each meeting. The interaction sheets were distributed to students 13 times during the semester. The weekly interaction sheets were coded using a grid for each week of the semester. These weekly grids were subsequently summarized into a single grid representing the total reported interactions among each pair of students during the semester. For a more detailed discussion of the methods used in this study, see Palmquist (1990, 1993).

Creating the Concept List. The concept list was established using an empirical approach. UNIX utilities were used to identify frequently occurring words and phrases in the interview transcripts, student journals, and transcriptions of the classroom observation notes. Terms and phrases found in at least 25% of the interviews or journal entries were added to the concept list. Terms and phrases that were used in at least three class sessions were also added to the concept list. In addition, key terms and phrases that were used in the textbook were added to the concept list.

Creating Maps. Maps were created for each student interview transcript and each student journal. A map was created for the instructor based on interviews, analysis of the textbook, and classroom observation notes. Each map was created using a two-step process. First, using the query/replace utility in a word processing program, terms and phrases in the concept list were marked in the text files (e.g., the term *write* was replaced with ___*write*). Second, the text files were printed and, using software in the MECA suite, maps were created for each text. A primary rater coded all texts analyzed in the study, and a second rater coded 10% of the interview transcripts and journal entries. Interrater reliability was calculated using signal-detection analysis (Lindsay & Norman, 1972). For the interview transcripts, interrater agreement was 95.5% for concepts and 70.7% for statements, with false alarms for the second rater of 6.5% and 38.6%, respectively. For the journal entries, interrater agreement was 93.1% for concepts and 65.5% for statements, with false alarms for the second rater of 5.8% and 33.8%, respectively.

Unlike the maps created in the robot study, no relationship types were defined in this study. Statements simply indicated that two concepts were related in some way. This procedure was followed primarily because this was an exploratory study of learning in an ill-defined domain. It seemed likely, based on previous research efforts, that students would provide relatively imprecise descriptions of their efforts to learn about writing. As a result, the choice was made to focus primarily on the use of concepts and the extent to which (although not the manner in which) those concepts were related.

Analyzing Maps. Analysis of the maps was conducted in a manner similar to analysis in the robot study. Maps were analyzed statistically and depicted graphically. A sample student map is depicted in Figure 10.5. Statistical measures of interest include the intersection of concepts (i.e., the number of concepts that any two maps have in common) between each student map and the instructor map and between each pair of student maps; the intersection of statements (i.e., the number of statements that any two maps have in common) between each student map and the instructor map and between each pair of student maps; the mean intersection of concepts between each student map and all other student maps; and the mean intersection of statements between each student map and all other student maps.

Results

The results presented in Table 10.3 indicate significant increases in a series of one-tailed tests in the intersection of concepts and statements between student maps and the instructor map and significant increases in the intersection of concepts and statements among student maps. The only test that did not indicate a significantly larger intersection was the nonsignificant increase (p = .07) for the intersection of concepts between student and instructor maps in the journal entries.

As might be expected, given the increases in the intersections of concepts and statements across student maps, several concepts and statements were

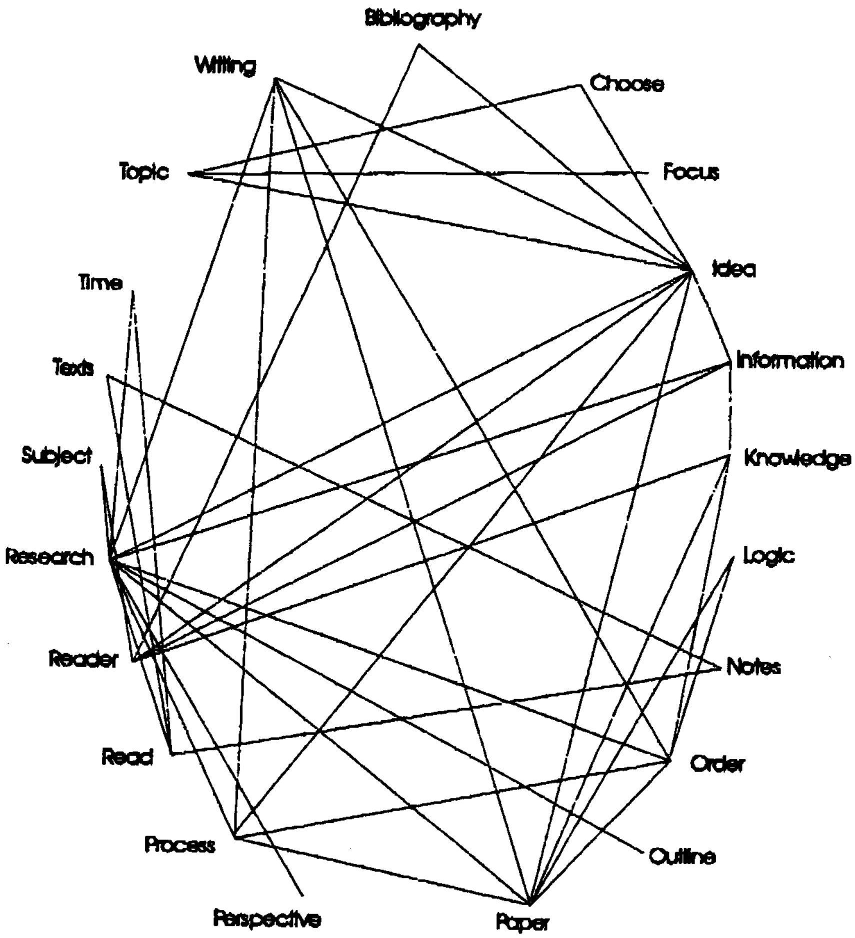

FIG. 10.5. Student map of first journal entry: "How and for what purpose(s) do you write a journal entry?"

used by more students at the end of the semester than at the beginning. As Table 10.4 indicates, concepts such as "analysis," "contribution," and "line of argument," which were used by less than 10% of the students in the first round of interviews, were used by over 60% of the students in their final interviews. Similarly, five statements found in at least 50% of the student maps based on the end-of-semester interviews were not present in any of the maps based on the first round of interviews. It is clear, however, that this growth in shared concepts and statements was not simply additive. Although a number of new concepts entered the lexicon of the classroom, several concepts that were used by a majority of students at the beginning of the semester saw substantial declines in use during interviews at the end of the semester (e.g., "outline," "organize," "notes," and "library"). Similarly, several statements were found in fewer maps at the end of the semester than at the beginning.

10. ANALY

C

Intersectio
Interviews
Journal ent
Interviews
Journal ent
Intersectio
Interviews
Journal ent
Interviews
Journal ent

Note. S
*p < .05

To det similariti pair of s concepts ter. Tota sented b the cells tion of s quadratic of two m score inc informati 1987.) In a procedur tion betw associati between

TABLE 10.3
Concept and Statement Intersections at Beginning Versus End of Semester

	Beginning of Semester	*End of Semester*	*t*
Intersections between student and instructor maps			
Concept intersections			
Interviews	13.1	27.4	5.82***
	(3.7)	(9.6)	
Journal entries	21.1	25.6	1.57
	(5.5)	(7.8)	
Statement intersections			
Interviews	5.8	40.3	5.19***
	(4.3)	(25.5)	
Journal entries	29.7	73.5	4.07***
	(16.5)	(32.3)	
Intersections between student and classmates' maps			
Concept intersections			
Interviews	11.6	19.5	7.09***
	(2.3)	(4.1)	
Journal entries	13.6	23.0	6.17***
	(2.2)	(4.8)	
Statement intersections			
Interviews	28.4	37.3	3.08**
	(6.0)	(12.3)	
Journal entries	25.9	45.5	4.44***
	(6.7)	(15.4)	

Note. Standard deviations are in parentheses below percents.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

To determine whether student interaction was positively associated with similarities among student maps, the number of interactions between each pair of students over the semester was compared with the number of concepts and statements each pair had in common at the end of the semester. Total interactions and shared concepts and statements were represented by data matrices in which the rows and columns were students and the cells were either total interactions per pair of students or the intersection of statements or concepts. The matrices were compared using the quadratic assignment procedure, which compares the corresponding cells of two matrices and, accounting for row and column effects, produces a z score indicating the probability that the two matrices covary. (For more information on use of the quadratic assignment procedure, see Hubert, 1987.) In addition to producing a probability score, the quadratic assignment procedure also indicates the direction (positive or negative) of the association between the matrices. It does not, however, indicate the strength of the association. Our analysis produced two significant positive associations between student interaction and shared concepts and statements (z = 1.85,

TABLE 10.4
Percentage of Concepts and Statements Mentioned in Interviews at Beginning Versus End of Semester

Concepts	*Semester*		*Statements*	*Semester*	
	Begin	*End*		*Begin*	*End*
organize	56	25	notes—research	69	13
outline	50	38	source—writing	63	13
knowledge	56	44	paper—topic	50	19
library	75	44	notes—paper	50	31
notes	88	44	notes—writing	69	31
source	75	44	research—source	75	31
argument	6	50	information—writing	75	38
books	69	50	library—research	69	38
solution	13	50	analysis—summary	0	50
subject	44	50	author—paper	0	50
talking	38	50	information—research	88	50
thinking	44	50	problem—writing	6	50
articles	19	56	author—summary	0	56
group	6	56	author—writing	19	56
issue	19	56	idea—writing	56	56
important	44	63	issue—research	6	56
line of argument	0	63	problem—research	13	56
topic	100	63	read—research	44	56
contribute	0	69	research—topic	94	56
difference	44	69	topic—writing	81	56
information	88	69	author—research	13	63
approach	13	75	contribute—research	0	63
newness	19	75	idea—research	56	63
point	44	75	newness—research	13	63
synthesis	13	75	research—summary	25	63
problem	19	81	analysis—synthesis	6	69
analysis	6	88	paper—writing	88	69
author	19	88	paper—research	94	75
find	75	88	summary—synthesis	0	75
read	75	88	research—writing	100	100
summary	31	88			
paper	94	94			
idea	69	100			
research	100	100			
writing	100	100			

$p < .05$, and $z = 1.89$, $p < .05$, respectively, in one-tailed tests). However, these associations were found only between out-of-class interactions and shared knowledge in the written descriptions of research writing. No significant associations were found between out-of-class interaction and shared knowledge in the student interviews. Nor were significant associations found between in-class interaction and either type of shared knowledge.

10. ANALY

Discussi

Result developn number creased, maps sha derived f behavior: to the da is additiv at least h in fewer

Methc potential in particu low (65% associate concepts accuratel

As we define rel ties betw relationsl maps of e might no becomes will be le require tl (Carley, 1 of these r

GENER/

The two cally anal concepts and statis approach and quan

Discussion

Results. This study illustrates how text analysis can be used to track the development of a shared lexicon. As the semester progressed, the mean number of concepts and statements shared between student maps increased, as did the mean number of concepts and statements that student maps shared with the instructor map. The study also illustrates how data derived from this form of text analysis may be associated with observable behaviors, in this case student interaction. Finally, the study calls attention to the dangers of assuming that growth in shared concepts and statements is additive. In this study, several concepts and statements that appeared in at least half of the student maps at the beginning of the semester appeared in fewer maps at the end of the semester.

Method. As was the case in the robot study, this study underscores potential problems associated with relying on human judgment. In this study in particular, the agreement between raters on statements in each text was low (65% to 70%). This study also calls attention to a loss of information associated with data reduction. Because the types of relations between the concepts in this study were not specified, it is impossible to reconstruct accurately the texts from which they were drawn.

As we noted previously, this study is exploratory. The choice not to define relationship types was a direct result of not being sure that similarities between texts would be found if these similarities required matching relationship types as well as concepts. Had it been possible to recode the maps of each text easily (in effect, to do a finer grained analysis), this choice might not have been made. As software for computer-aided text analysis becomes more sophisticated, it seems likely that the tradeoffs in this study will be less common. However, for the time being, logistic constraints often require the text analyst to make a large number of self-restricting choices (Carley, 1993). Software developments in the near future may alleviate many of these restrictions (Bechtel, this volume).

GENERAL DISCUSSION

The two studies in this chapter demonstrate how researchers can empirically analyze shifts in meaning, definition, and intent by focusing on both concepts and the relationships among concepts. In both studies graphical and statistical devices are used to address research issues. In this way, the approach allows the researcher to move back and forth between qualitative and quantitative analysis.

Currently, it is quite difficult to analyze maps. Indeed, our ability to code complex maps far outstrips our ability to compare and analyze these maps. Complex maps may have multiple types of relations, uni- and bidirectional relations, and variable strengths on the relations (e.g., Kleinnijenhuis, de Ridder, & Rietberg, this volume). Such complexity differentiates and gives power to the coding scheme. However, current statistical techniques for analyzing maps are most reliable and most highly developed for simple maps. Future research needs to address the statistical comparison of more complex maps.

The simplest map has all relations of the same type—relations that are either uni- or bidirectional and that are either present or not. Maps coded in this way can be described as binary graphs. Comparison and analysis techniques for binary graphs are fully developed. For maps coded in this way, the hamming metric can be used for determining the extent of similarity. For a set of such maps, it is possible to locate the central map (as was done in the writing study) and to estimate its sampling distribution (Banks & Carley, 1994). This is comparable to locating the mean and standard deviation for a variable. These statistics can be used to test hypotheses about differences among classes of maps. For example, with respect to the robot study, these statistics make it possible to test whether the cultural perception of robots has shifted significantly over time. The significant trend from negative to positive attitudes toward robots was evaluated in analyses of single statements. The Banks and Carley technique would allow the researcher to simultaneously focus on all statements in each map.

Aside from the methodological issue that statistical techniques for complex graphs are limited, there is also a theoretical issue: What theoretic purpose does this greater complexity serve? Clearly, with more complex coding schemes the ability to regenerate text from code is increased (Roberts, this volume, chapter 3). Further, such complexity may have diagnostic value at the individual level (Gottschalk, this volume). However, the greater the complexity of the coding scheme, the lower the likelihood that two maps will appear similar. This issue was described earlier as a tradeoff between validity and data reduction. Specifically, the more data are reduced, the more coder agreement is enhanced, but, quite likely, the less valid will be one's encoded data and the less likely these data can be accurately reconstructed into a semblance of the original text. The rule of thumb seems to be that the researcher should only opt for the degree of complexity required to answer the research question at hand.

There is a final cautionary note, however. Interrater reliability varies negatively with the complexity of the coding task. We found great differences among coders in their willingness and ability to make inferences and to generalize. Moreover, coders are more similar in their treatment of concepts than they are in their treatment of statements. This suggests that applications of map analysis must tailor coder training to increase interrater reliability with respect to statements. We see three possible avenues worth exploring: development of a (possibly hierarchical) theory of concept relations (cf. Roberts, this volume, chapter 3), development of computer-aided procedures that decrease reliance on human coders (cf. Popping, this volume), and further development of tools like SKI that make explicit implicit relations and so augment human coding by focusing on the missed relations.

The two studies we have presented illustrate the power of using maps to analyze and compare mental models. Even simple maps admit empirical analysis and make it possible to address detailed questions about shifts in meaning. Further, map analysis allows us to address learning in terms of the structure and the content of what is learned. We can ask, with these maps, do mental models become more complex or more integrated over time? Under what conditions do novice mental models become more like expert models? What is the content of this change? With map analysis, we can thus focus not only on concept usage but also on meaning and shifts in cognitive structure.