CHAPTER

4

NETWORK TEXT ANALYSIS: THE NETWORK POSITION OF CONCEPTS

Kathleen M. Carley
Carnegie Mellon University

Mental models can be abstracted from texts as a network of relations between concepts. Within these mental models different concepts play different roles depending on their position in the network. This chapter describes how to locate and empirically describe each concept's position in a mental model and how to construct composite empirical descriptions of the entire mental model. The measures and techniques for comparing texts using this framework are illustrated using data gathered from a study of undergraduates selecting a new tutor for their living groups.

Language is a chronicle of social knowledge that is predicated on the society's history, culture, and social structure (Cicourel, 1970; Vygotsky, 1962; White, 1992). Language contains within it societal choices about how to represent and interrelate concepts. Such choices frame the way individuals think about the world and so affect what actions individuals take. Thus, language affects behavior (cf., e.g., Cicourel, 1974; Cooley, 1902; Stryker, 1980). Language can be represented as a network of concepts and the relationships among them (Axelrod, 1976; Schank & Abelson, 1977; Sowa, 1984). This network can be thought of as the social structure of language or, equivalently, the representation of extant social knowledge. The social structure of language is related to the structure of action because these societal choices are embodied in the linguistic social structure.

Multiple techniques exist for extracting conceptual networks (cf. Carley, 1988; Carley & Palmquist, 1992; Kleinnijenhuis, de Ridder, & Rietberg, this volume). Such networks can be empirically analyzed, and the analysis may

LEA—MARTIS GRAPHICS—TEXT ANALYSIS FOR THE SOCIAL SCIENCES (ROBERTS)

shed light on individual actions (Carley, 1986a; Carley & Kaufer, 1993; Kaufer & Carley, 1993b; Palmquist, 1990), political discourse (Kleinnijenhuis, de Ridder, & Rietberg, this volume), narratives, or scripts (Abell, 1984; Heise, 1991). This chapter builds on this prior work. I argue that:

- Language can be represented as a lossy-integrated conceptual network.[1]
- Each concept has a position in this network that can be characterized along several dimensions.
- Conceptual networks can be characterized by the distribution of concepts on this set of dimensions.
- Analysis of a society's language using these dimensions may provide insight into the use of language by members of the society to build consensus.

A network-based procedure for analyzing conceptual networks is presented. The procedure enables the researcher to characterize the structure of the conceptual network along a series of dimensions. These dimensions are independent of how the network was extracted and the source from which the network was extracted. The interpretation of the results (i.e., how concepts cluster along these dimensions), however, is dependent on both the method of extraction and the source of the network.

The proposed procedure for analyzing conceptual networks allows the researcher to locate the position of each concept in the conceptual network, classify concepts according to a taxonomy, and make predictions about action on the basis of the distribution of concepts in this taxonomy. The proposed procedure is then used to analyze the talk of a group of undergraduates at MIT. The results are used to illustrate the link between the social structure of language and action through an examination of the power of concepts to evoke consensus. The MIT data set includes information on the social language (i.e., social knowledge) and individual language about the topic of tutor (i.e., graduate resident) selection for a group of undergraduates in a single living group (Carley, 1984, 1986a).

This study is exploratory. The goal is not to explain the language of this particular group. Rather, the goal is to present a model of language as a social phenomenon and a procedure for using this model to look at the interrelations

[1]Lossy integration refers to an integration process in which information is lost, as when taking a moving average. A lossy-integrated network is one in which the relations between nodes change gradually over time through a moving average process. For language, the basic idea is that each person's language can be represented as a conceptual network. Social language can be represented as a composite network by combining (somehow) the networks of all individuals in the society at that point in time. Because individuals enter and leave societies, and because individuals can learn and evolve their personal networks, the network representing social language is lossy.

between language and action. The MIT talk is used purely as an illustration. Previously, the MIT talk was coded as a specific type of conceptual network—the cognitive network or mental model (Carley, 1984, 1986a). In such networks the meaning of each concept is captured by its network position (Carley & Palmquist, 1992; Carley, 1993; also see Fauconnier, 1985, and Johnson-Laird, 1983, for theoretical underpinnings). Using the proposed procedure these cognitive networks can be analyzed. However, the proposed procedure can also be applied to other types of networks. For example, this procedure could be applied to scripts or narratives where the relations between concepts represent story order or event sequences (Abell, 1984; Heise, 1991).

DEFINITIONS

A conceptual network is characterized as a set of concepts and pairwise relations between them. Numerous schemes for representing conceptual networks exist in the literature. These schemes vary in the semantic grammars used for ascribing a relation among concepts (Franzosi, 1990a, this volume; Kleinnijenhuis, de Ridder, & Rietberg, this volume; Roberts, 1989, this volume). The specific scheme used in this chapter is grounded in map analysis (Carley, 1984, 1986a, 1993; Carley & Palmquist, 1992). Concepts, relationships, and statements are the basic network components of conceptual networks.

A *concept* is an ideational kernel—a single idea. Graphically a concept is represented by a circle (in Figure 4.1, for example, "Aria," "Cassi," "friends," "gives," "plays," and "toy" are all concepts). In cognitive networks (both at the individual and societal, or socioconceptual, levels), concepts are devoid of meaning except as they relate to other concepts.

A *relationship* links two concepts. Graphically a relationship is represented by a line connecting two circles. In Figure 4.1, the relationships are labeled as (a) through (i). Relationships can have a variety of properties, of which the only two that are important to this study are strength and directionality.[2] In sociocognitive networks the strength of a relationship indicates the degree of consensus—the extent to which members of the population agree that there is a relationship between these two concepts. In cognitive networks the

[2]Researchers have also included relationship type and sign as properties of relationships (see Carley, 1984, 1986a, 1988; Franzosi, this volume; Kleinnijenhuis et al., this volume; Roberts, this volume). In this chapter, all relationships are treated as of the same type (namely, a→b and/or b→a). If the researcher has multiple types, the procedures outlined in this chapter can be followed for each type of relationship separately. Although many approaches to coding texts allow for relationships to be both negative and positive, such an approach is not suggested when dealing with conceptual networks. The basic reason is that negated concepts may erroneously be interpreted as implying the opposite of positive concepts (e.g., that "not in love" has the opposite meaning of "in love"). To avoid making this assumption, which generally does not appear to hold, it is recommended that the researcher when generating the conceptual network use only positive relationships and separate positive from negative concepts.

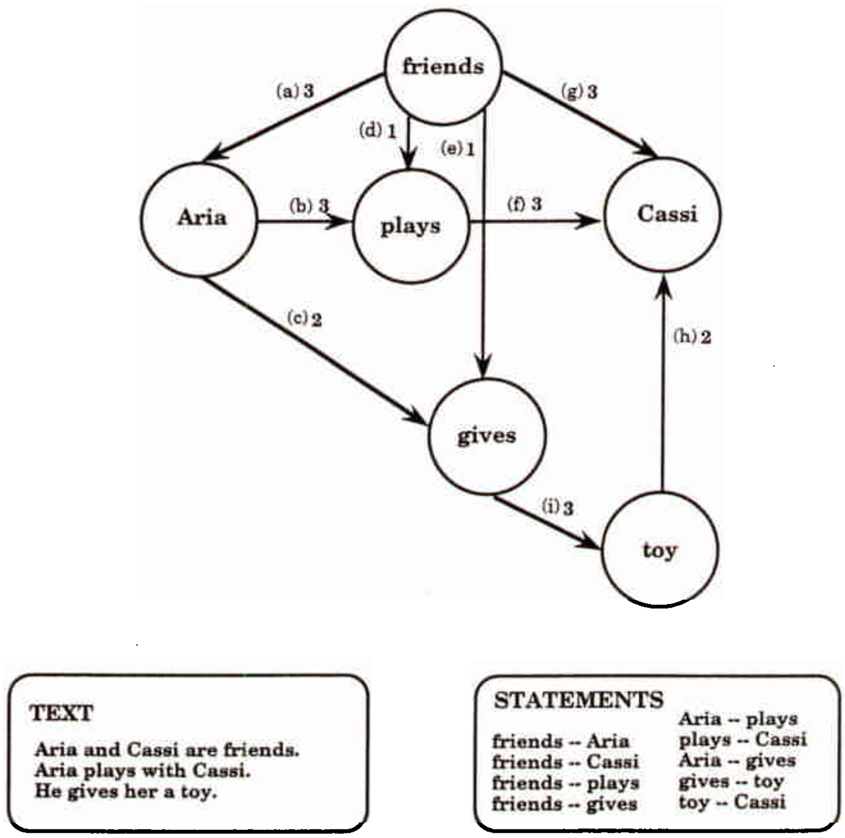

FIG. 4.1. Illustrative conceptual network.

strength of a relationship indicates the degree of emphasis, salience, implication, or belief—the extent to which the individual emphasizes or believes that the relationship between the two concepts holds. In Figure 4.1, for example, the strength of the relationship denotes the degree of inference that must be made to link the two concepts such that a "3" represents those relationships that are directly stated, "2" those relationships that can be inferred from syntax, and "1" those relationships that can only be inferred using tacit social knowledge not explicit in the text (cf. Carley, 1988).

A *statement* is two concepts and the relationship between them. Because relationships are directed, within a statement one concept is in the anterior position and one is in the posterior position. For example, in Figure 4.1, in the "Aria→plays" statement, "Aria" is in the anterior position, and "plays" is in the posterior position.

In conceptual networks, the relative network position of each concept can be measured along several dimensions. To define these dimensions six additional ideas are needed. *Vocabulary* includes the set of concepts in the conceptual network. To facilitate later discussions, let us represent the number of concepts in the conceptual network as N. The size of the vocabulary is N. The *focal concept* is the concept whose network position is being measured. To characterize the conceptual network completely, each concept in the vocabulary is in turn treated as the focal concept. Two concepts that occur in a single statement are said to be *directly linked* to each other. In graph representation, a concept is directly linked to those concepts to which it is linked by an arrow; "Aria" and "friends" are directly linked in Figure 4.1. An *indirect link* exists when two concepts do not occur in the same statement but are linked by a directed chain of statements. In graph representation, two concepts are indirectly linked when a path (following the arrows) exists between the two concepts with at least one intervening concept. In Figure 4.1, "Aria" is directly linked to "gives" and indirectly linked to "toy."

For a focal concept, its *local network* is the set of concepts to which it is directly linked. In Figure 4.1, the local network when "Aria" is treated as the focal concept includes the concepts "friends," "plays," and "gives" and the relationships (a) through (c). For cognitive networks, a concept's local network is the locally elaborated meaning of that concept. The *extended network* for a focal concept can be generated for each concept in the larger network by following the procedure illustrated in Figure 4.2 and described here. For cognitive networks, a concept's extended network defines the generative meaning of the concept.

In generating the extended network, relationship strengths are used. First, a cutoff is defined.[3] Those statements whose relationships have a strength greater than the cutoff are treated as definitives, defining what other concepts must be present. Those statements whose strength is less than the cutoff are treated as connectives, defining for co-present concepts what relationships must exist.

The process of generating an extended network for the focal concept uses a forward-chaining statement-inclusion method. The procedure begins with the focal concept. Then, using those definitives that have as their anterior concept the focal concept, a set of concepts (and the relationships) are added. This process is repeated, recursively, until no new concepts can be added. Then the connectives are used to fill in relationships between concepts in the extended network. In Figure 4.2, this procedure is followed for each of the concepts in the conceptual network in Figure 4.1, using the average strength as the cutoff. Not all concepts have an extended network (e.g., "Cassi" and "toy"). In cognitive networks, the extended network can be thought of as the

[3]The researcher can define the cutoff using any criteria. In the program CUBE, the default for the cutoff is the average strength of all relationships in the conceptual network. CUBE is written in C and can be obtained from the author. It is part of the MECA software package for encoding and analyzing maps.

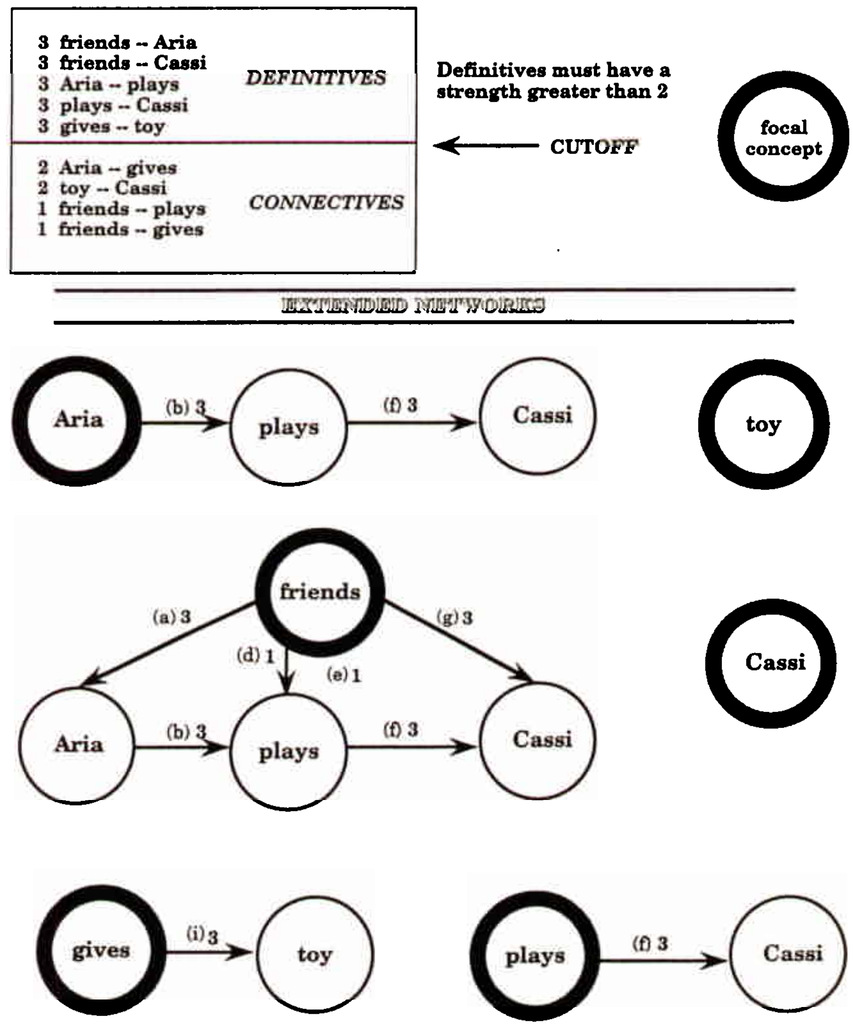

FIG. 4.2. Generating the extended network.

image that is most emphasized (in the individual's mental model) or most likely to be evoked (across all individuals in the society when the network is the sociocognitive network). The extended network provides insight into the train of thought and not just the direct inferences.

DIMENSIONS

A concept's position in the conceptual network can be characterized along five connective dimensions: imageability, evokability, density, conductivity,

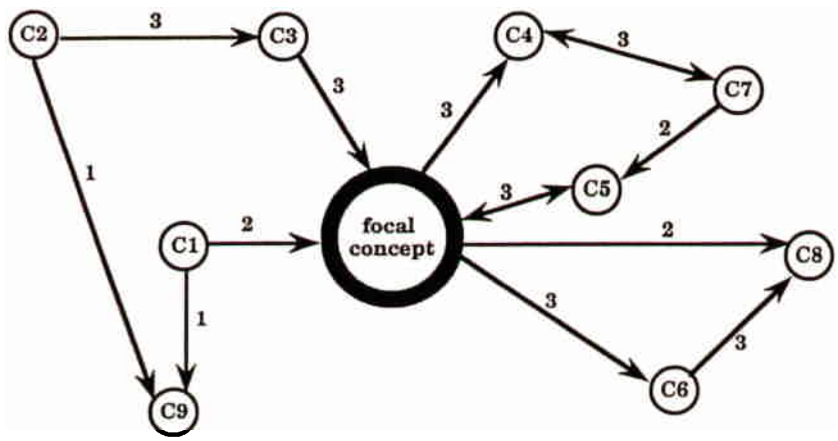

FIG. 4.3. Connective properties of concepts.

and intensity.[4] These dimensions have analogues at both the local level and the extended level. The dimensions can be thought of as measuring the connective properties of the concept, that is, as measuring the nature of each concept's connection to other concepts. In defining each of these dimensions, the abstract conceptual network in Figure 4.3 can be used. Each of these dimensions has a theoretical maximum. When analyzing data it is sometimes useful to consider a concept's absolute value on a dimension and at other times to consider its value as a percentage of the theoretical maximum. In Figure 4.3 the strength of the relationships are noted as 1,2,3. A strength of 3 denotes a definitive.

Local imageability is measured as the total number of statements in the map that contain focal concepts in the anterior position. Graphically, this is the number of arrows going from the focal concept to other concepts. The local imageability of the focal concept in Figure 4.3 is 4. The theoretical maximum is N–1, as the concept can have at most a relationship to all concepts other than itself.

Local evokability is measured as the total number of statements in the map that contain the focal concept in the posterior position. Graphically, this is the number of arrows going into the focal concept from other concepts. The local evokability of the focal concept in Figure 4.3 is 3. The theoretical maximum is N–1, as the concept can have at most a relationship from all concepts other than itself.

Local density is measured as the total number of statements in the map that contain the focal concept in either the anterior or the posterior position.

[4]The program CUBE calculates, for each concept in the social vocabulary, the connective properties of that concept within a conceptual network coded as a map. The output from CUBE is a matrix of concepts by their position on each of the connective dimensions.

Local density is operationalized as the sum of local imageability and local evokability. The local density of the focal concept in Figure 4.3 is 7. The theoretical maximum is 2*(N−1).

Local conductivity is the number of two-step paths through the focal concept. Local conductivity is operationalized as local imageability times local evokability. The local conductivity of the focal concept in Figure 4.3 is 12. The theoretical maximum is $(N-1)^2$.

Local intensity is the strength of the focal concept's direct relationships to other concepts. This is measured as the fraction of the statements that contain the focal concept in either the anterior or the posterior position with greater than average strength. The local intensity of the focal concept in Figure 4.3 is 0.71 (5/7). The theoretical maximum for local intensity is 1.

Extended imageability is measured as the total number of concepts in the focal concept's extended network. The extended imageability of the focal concept in Figure 4.3 is 5. The theoretical maximum is N−1, as at most all other concepts can occur in the extended network.

Extended evokability is measured as the total number of concepts in whose extended network the focal concept occurs. The extended evokability of the focal concept in Figure 4.3 is 3. The theoretical maximum is N−1, as at most the focal concept can occur in all other concept's extended networks.

Extended density is measured as the total number of concepts that either occur in the focal concept's extended network or in whose extended network the focal concept occurs. Extended density is operationalized as the sum of extended imageability and extended evokability. The extended density of the focal concept in Figure 4.3 is 8. The theoretical maximum is 2*(N−1).

Extended conductivity is the number of pairs of concepts, such that neither of the two concepts are the focal concept, that are linked by at least one path through the focal concept. Extended conductivity is operationalized as extended imageability times extended evokability. The extended conductivity of the focal concept in Figure 4.3 is 15. The theoretical maximum is $(N-1)^2$.

Extended intensity is the strength of the focal concept's extended network. This is measured as the fraction of concepts that either occur in the focal concept's extended network or in whose extended network the focal concept occurs and that are strongly tied to the focal concept. The extended intensity of the focal concept in Figure 4.3 is 0.63 (5/8). The theoretical maximum for extended intensity is 1.

Let us pause and consider the conceptual bases of the three basic dimensions, density, conductivity, and intensity. Each of these dimensions can be thought of as measuring an aspect of the concept's communicative power. The communicative power of a concept is a multidimensional notion involving the extent of the consensus or shared meaning that occurs when a concept is used and the extent to which the use of the concept can affect social change. Concepts that are high in density derive communicative power from the wealth of other concepts to which they are attached. Highly dense concepts are likely to be used and thought about. Concepts that are high in conductivity derive their communicative power from their ability to tie groups of (otherwise largely disconnected) concepts together. Highly intense concepts derive their communicative power from the degree to which there is social consensus over their relations to other concepts.

A GENERAL TAXONOMY

Using the dimensions of density, conductivity, and intensity, a taxonomy of concepts can be formulated that derives its power as a classification scheme by simultaneously "typing" concepts and providing a framework within which the evolution of concepts, and hence knowledge, relative to a specific task can be analyzed. This classification scheme categorizes concepts according to the potential role concepts play in communication. The identified types or classes of concepts are in effect ideal types, outliers whose position vis-à-vis the three dimensions defines the cube. Each type of concept has particular communication properties and may even have a characteristic label relative to the type of conceptual network. To facilitate the discussion of the data, these eight ideal types are labeled and described as though the conceptual network were a cognitive or sociocognitive network. The eight ideal types are ordinary concepts, prototypes, buzzwords, factoids, place-holders, stereotypes, emblems, and symbols. Within cognitive and sociocognitive networks, if language is a lossy-integrated network, then the utilization of concepts should change as their level of density, conductivity, and intensity changes.

One way of looking at this taxonomy is as differentiating concepts in terms of their embedded meaning. The network of concepts evoked by the focal concept can be thought of as its embedded meaning. Concepts with higher levels of embedded meaning are going to be more dense. Concepts with a more temporal meaning (i.e., likely to be evoked and to evoke) are going to be more conductive. Concepts with more historically developed networks (i.e., the network of concepts has arisen in response to historical events) are going to be more intense (Carley & Kaufer, 1993; Kaufer & Carley, 1993a, 1993b).

Communicative Power and Network Position

One way of seeing the potential communicative power of concepts is to consider their relative position in this classification scheme. Let us consider the theoretical implications of a concept being extreme on one or more of these three dimensions.

Type 1: Ordinary concepts. Ordinary concepts are low on all three dimensions. Such concepts are isolated concepts within a system or network of concepts. Most concepts should be low on all three attributes. Ordinary concepts are used to define the critical or outlying concepts but are in themselves (relative to the discourse topic) of little importance. The use of ordinary concepts should neither generate nor inhibit consensus formation.

Type 2: Prototypes. Prototypes are high in density but low in conductivity and intensity. Prototypes are characterized by being connectively central but nevertheless perhaps of little import. Prototypes have an elaborate meaning that is not highly emphasized or agreed to. Such concepts should be handles for a set of ideas whose meaning (i.e., set of relationships to other concepts) has been affirmed historically but not to the extent that there is widespread consensus as to the existence or interpretation of these relationships. The movement from ordinary concepts to prototypes is a movement from the general and astructural to the historical and negotiable.

Type 3: Buzzwords. Concepts that are high in conductivity but low in both intensity and density are defined as buzzwords. Buzzwords are concepts that, relative to a particular topic, are highly utilized by individuals when discussing that topic. Such concepts have little meaning, and what meaning they do have is not particularly salient or socially shared. In a very loose sense, buzzwords are the result of individual belief or social consensus that a single idea should be important and highly relevant to the task at hand. The movement from ordinary concepts to buzzword is a movement from concepts that are very general and whose meaning is more a function of personal experience (i.e., little social or shared meaning) to concepts that are highly temporal and vogue. Buzzwords are very astructural in meaning (i.e., they have none) and are part of the more transient popular culture and not the stable long-term underlying culture. The movement from ordinary concepts to buzzwords is a movement from concepts that are atemporal to those that are temporal.

Type 4: Factoids. Factoids are defined as concepts that are high in intensity but low in density and conductivity. Such concepts have a narrowly ascribed meaning that is nonetheless highly salient or accepted. Within a general discourse community, such concepts are likely to be culturally shared identifications or trivia, such as dates: "1492," "1776." Trivia games rely on such concepts. Within a particular type of discourse community or a profession, such concepts are likely to be agreed upon definitions such as sociometric and demographic terms (e.g., age, sex). Such concepts underlie and serve to define many other concepts; however, as isolated concepts they have low levels of usefulness relative to any particular task. The movement from ordinary concepts to factoids is a movement from concepts with individualized meanings to concepts whose meaning is culturally embedded.

Type 5: Place-holders. Place-holders are defined as high in density and conductivity and low in social consensus. Such concepts should be highly utilized. Place-holders, like buzzwords, admit the construction of consensus by producing a situation in which there is tacit consensus to an ill-defined entity, an entity whose formulation has not been socially consented to. Unlike buzzwords, place-holders serve as handles for a large set of ideas that bear some relationship to each other and are highly relevant to the task at hand. Place-holders and buzzwords should facilitate communication and pave the way for consensus formation and the resultant evolution of sociocultural knowledge. Their high level of usage implies tacit agreement that "x" should be important without social agreement as to what "x" means. The movement from ordinary concepts to place-holders is a movement from general meanings to highly task-relevant meanings. Place-holders have meaning; it's just highly negotiable.

Type 6: Stereotypes. Stereotypes are defined as concepts that are high in intensity and density but low in conductivity. Stereotypes represent historical saliency or consensus to regularities perceived by members of the social unit. They should be highly structured images. Stereotypes should change slowly due to the low level of conductivity. Stereotypes can be thought of as highly consented-to prototypes. The movement from ordinary concept to stereotype is a movement from the astructural to the structural.

Type 7: Emblems. Emblems are defined as concepts with high intensity and conductivity and low density. Emblems are concepts that are highly utilized relative to the task at hand but that have a very narrow, highly consented-to meaning. Emblems are useful communication tools as they admit instant identification. The movement from ordinary concepts to emblems is a movement from concepts whose meaning is consented to only by the individual to concepts whose meaning is socially consented to.

Type 8: Symbols. Symbols are defined as the sociocultural antithesis of ordinary concepts. Whereas ordinary concepts are low in all dimensions, symbols are high in all dimensions. One type of concept that acts as a symbol would be social roles. To the extent that this taxonomy holds, it should be possible to measure the level to which a particular role has become an accepted aspect of that social unit by measuring the degree to which the concept denoting that role is high in all three dimensions (either across all individual's cognitive networks or within the overall sociocognitive network). The movement from ordinary concepts to symbols is a movement from concepts with very general purpose and highly personal meaning and that are very astructural to concepts that are highly relevant to the task at hand, have strong social meanings, and are highly structured. The movement from ordinary concepts to symbols is a movement from a single conceptual entity to a sociocultural construct whose conceptual handle is relevant and highly embedded.

ILLUSTRATIVE EXAMPLE

Data Collection and Coding

The data were collected as part of a larger study on consensus construction (Carley, 1984) in which the language, social structure, and decision-making behavior of a group of undergraduates (all members of the same living group) were studied as they went through a process of tutor selection. For an overview of this study, refer to Carley (1986a). In this chapter, the concern is with the language used by the students to talk about tutors and tutor selection and how that language can be used to induce consensus.

As they went through this process, a variety of verbal information on what the students wanted in a tutor, how they thought about tutors, what they thought tutors did, the social history of tutors in that living group, and so on, was collected. Based on these data the language used by the students to talk about tutors was coded (Carley, 1988). These language data do not reflect the entire sociocultural environment and language of this group but only that part of their culture and language that relates to the tutor and the tutor selection process. Only a portion of the language data is used in this chapter—the general sociocognitive network. This illustrative network is coded from guided freeform interviews with four students (5% of the original sample) as well as from the sociohistorical records of the living group. (See Carley, 1984, Appendix 3, for more detail on these records.)

Social Vocabulary. The social vocabulary contains 210 concepts. Some concepts are single concepts (e.g., "gnerd"), whereas others are phrases (e.g., "fits in with hall"). Concepts are nouns or noun-based phrases. In most cases, the concepts are those actually used by the students. In some cases, however, they are paraphrases. The vocabulary is not complete in terms of the actual concepts used by the students. It is complete in terms of the set of general concepts needed to represent the perceptions about tutors forwarded by current and past residents of the living group. Roughly, this set of 210 concepts can be thought of as the set of concepts needed to engage in "tutor-talk."

Socioconceptual Network. In coding the sociohistorical records and interviews, each sentence, clause, or paragraph that contained tutor talk was coded as a statement using CODEMAP.[5] Details on the process of coding the socioconceptual network can be found in Carley (1988). Essentially, if two concepts occurred in the same sentence, for example, as subject and object, or if the two concepts logically followed one from the other within a paragraph, then a relationship was placed between those two concepts, from the first to the second, resulting in a single statement. The strength of each statement reflects the degree of agreement or consensus among members of the previous and current members of the living group that the two concepts are related and that when the anterior concept is used, the posterior is implied. Carley (1988) defined three strength levels (from high to low): definitives, logical connectives, and simple connectives. Definitives are statements where one concept defines the other, such that in the society in question, if the first concept is used the second is always implied. Logicals are statements where the concepts are logically connected, such that in the society in question, if the first and second concept are used the speaker intends a specific relation among them. Simple connectives are statements such that in the society in question, if the two concepts are used and the speaker has not specified an alternative relation, then the socially accepted relation between them is assumed. The socioconceptual network contains 1,214 statements, of which 275 are definitives, 837 are logical connectives, and 103 are simple connectives. Typical statements in this network are "someone who water fights encourages a fraternal atmosphere" and "someone who is older is lookupableto." In effect, the socioconceptual network can be loosely thought of as the set of sentences that have been and are commonly used by the students to conduct discussions about tutors. The result of the coding process is a single network with 210 concepts and 1,214 relationships among them.

Statistics and Distributions

For the socioconceptual network, univariate statistics for each dimension are shown in Table 4.1. These statistics are based on treating each of the 210 concepts in the socioconceptual network in turn as the focal concept. On average each of the 210 concepts connects to (imageability) and is connected to (evokability) 6 other concepts, is involved in 12 statements (density), is the center of 42 paths (conductivity), and has an average intensity of each relationship of 1 (moderate). The average profile is far from the theoretical maximum (and the adjusted theoretical maximum). In other words, the socioconceptual network is very sparse.

[5]CODEMAP is an extended version of CODEF, which is described in Carley (1988). CODEMAP is part of the MECA software.

TABLE 4.1
Univariate Statistics for Sociocognitive Network

	Mean	*Standard Deviation*	*Skew*	*Actual Maximum*	*Theoretical Maximum*
Local network					
Imageability	5.78	7.46	3.39	52	209
Evokability	5.78	5.52	2.72	44	209
Density	11.56	10.17	2.41	67	418
Conductivity	41.99	90.69	5.43	780	43,681
Intensity	0.92	0.14	−3.50	1	1
Extended network					
Imageability	3.38	4.00	1.65	19	209
Evokability	3.38	8.03	4.21	66	209
Density	6.76	8.74	3.06	66	418
Conductivity	8.74	28.63	5.14	180	43,681
Intensity	0.47	0.35	0.30	1	1

Note. Averages are across the 210 concepts in the network, using each in turn as the focal concept.

In Figure 4.4 the distribution of tutor talk concepts across local density, local conductivity, and local intensity is shown. Concepts tend to cluster on all three dimensions. Within this community, most words are either highly imageable or highly evocable but are rarely both. There are concepts that do stand out on one or more dimensions and therefore hold a special place in the proposed taxonomy. These concepts have special communicative properties.

Application of the Taxonomy

We would expect that were we to analyze all concepts in the English language for the American public that most would be ordinary concepts; that is, low on the three dimensions. As such, most concepts should play a fairly similar albeit unimportant role in terms of communication and consensus formation. On any one dimension, we would expect only a few concepts to stand out. For language about a specific topic, however, we might expect a different distribution, particularly given that researchers coding that talk may tend not to code most ordinary words such as *a, an,* and *the.* Nonetheless, the particular distribution of concepts for that talk should provide insight into the nature of the arguments being put forth by the speakers. Let us turn now to an examination of the MIT tutor talk.

Concepts in the socioconceptual network were classified into their requisite category using the following cutoffs. A density greater than or equal to 24 was considered high, a conductivity greater than or equal to 572 was considered high, and an intensity greater than 0.96 was considered high.

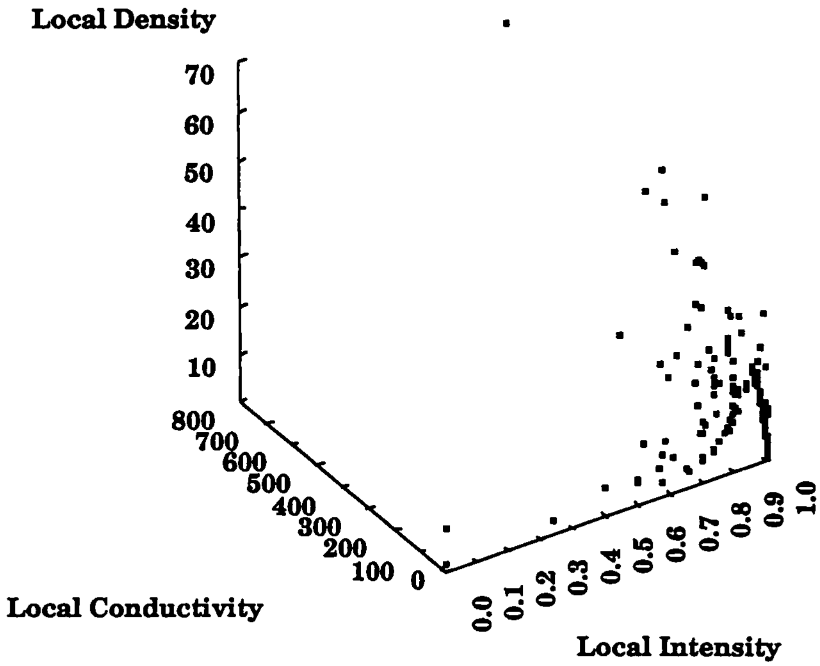

FIG. 4.4. 3-D projection of density, conductivity, and intensity at the local level.

Density and conductivity have a potential range based on the total number of concepts. However, most conceptual networks are quite sparse (i.e., very few concepts are connected to each other). Therefore, it makes sense to adjust behavior based on the achieved maximum. The cutoffs chosen here are .25 of adjusted maximums. For density the adjusted maximum is the sum of the actual maximums for imageability and evokability. For conductivity the adjusted maximum is the product of the actual maximums for imageability and evokability. These adjusted maximums are based on the assumption that potentially every concept can be connected as much as the most connected concept.

Given these cutoffs, most concepts in this socioconceptual network are either factoids or ordinary concepts (see Table 4.2). The high number of factoids is not particularly surprising given that this network represents knowledge used in making a decision. Factoids, which have a narrow but accepted meaning, allow individuals rapidly to describe similarities and differences between those items they are trying to decide upon. As such, factoids play a highly specialized but important communication role in decision making processes.

The concept "car" is a typical factoid. In Figure 4.5, the local network for the focal concept "car" is shown. The arrows show the direction of implica-

TABLE 4.2
Number of Concepts per Category

Category	*D*	*T*	*I*	*Examples*	*N*
ordinary concepts	L	L	L	finishing thesis	59
				humorous answer	
prototypes	H	L	L	intelligence	5
				lookupableto	
buzzwords	L	H	L	friendly*	0
				mature*	
factoids	L	L	H	car	133
				bar hops	
place holders	H	H	L	fits in	2
				gets along	
stereotypes	H	L	H	gnerd	11
				hacker	
emblems	L	H	H	BS from MIT*	0
symbols	H	H	H	ex third easter*	0

Note. Asterisks identify close examples that are near but do not fall into the corresponding cell using the specified cutoffs.

tion. The number on the line indicates the strength of the relationship. Whether the relationship is positive or negative is not indicated. Notice that this concept is rarely evoked and has a small image. However, there is high agreement. In particular, there is unanimous agreement that if the candidate has a car then he or she will encourage hall activities and that a car is a resource. If the students want to judge the candidates' ability to promote social interaction or encourage hall activities, they may well ask (and they did) whether the candidate has a car. If the candidate has a car, the students will tend to assume that the candidate will encourage activities within the living group. Factoids, given this small but highly consensual meaning, can be used in decision making scenarios to narrow the field of choices quickly.

Consider the concept "hacker." The concept hacker is classified as a stereotype. In Figure 4.6, the local network for the focal concept "hacker" is shown. The arrows show the direction of implication. The number on the line indicates the strength of the relationship. Whether the relationship is positive or negative is not indicated. There are many things that the individual can do to be classified as a hacker and many things that a hacker will do. However, there is greater agreement on when an individual is a hacker than on what a hacker will do. For example, all students agree that someone who is interesting, sits in the lounge, and tells stories is a hacker, but not all students agree that hackers are interesting. In other words, among these students, being interesting is a sufficient but not a necessary condition for being a hacker. Consensus about whether a candidate is a hacker may be quickly achieved, but whether that translates into a good

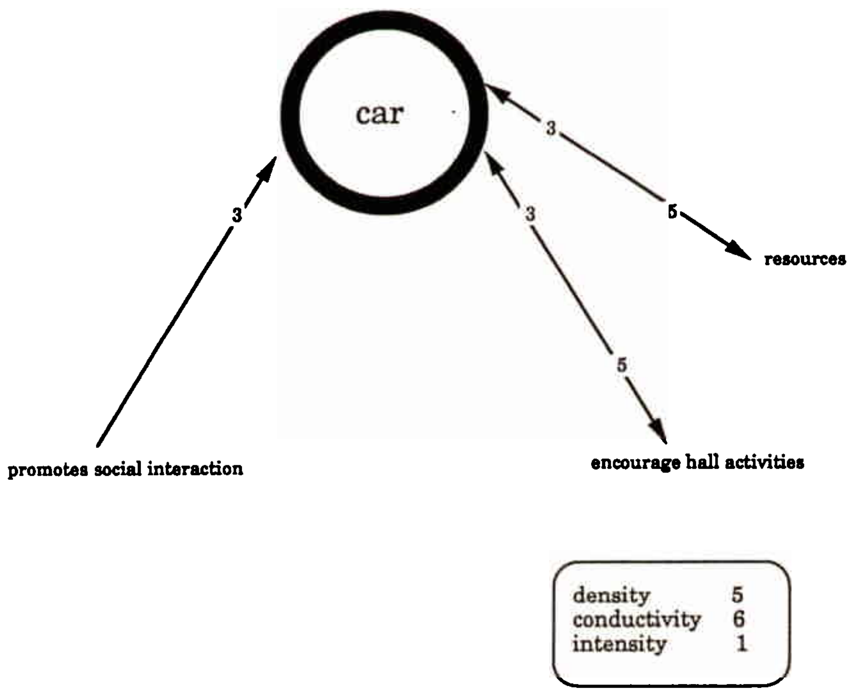

FIG. 4.5. Local network for the concept "car."

tutor candidate is ambiguous. Students in the living group may recognize only the agreement that the candidate is a hacker and so presume that there is agreement about the candidate as tutor, thus generating a false sense of consensus.

Consider the concept "gnerd." Gnerd is also classified as a stereotype; however, relative to hacker it is much closer to being a symbol. In Figure 4.7, the local network for the focal concept "gnerd" is shown. The arrows show the direction of implication. The number on the line indicates the strength of the relationship. Whether the relationship is positive or negative is not indicated. The concept "gnerd" almost serves as a symbol for this community as it has high density, conductivity, and intensity. Individuals labeled as gnerds play highly specialized social roles.

Labeling someone as a gnerd can have a particularly powerful negative effect on their chances of being a tutor, as "gnerd" is linked to two other stereotypes, "hacker" and "jerk." The negative link to hacker means that if the candidate is a gnerd then he or she is not a hacker. Gnerds are also considered jerks by some students. Thus, if a candidate is labeled as a gnerd, there can be strong ripples throughout individuals' cognitive models; however, these ripples have great variance across the community. "Hacker" produces similar ripples, but there is less variance in the ripples that are

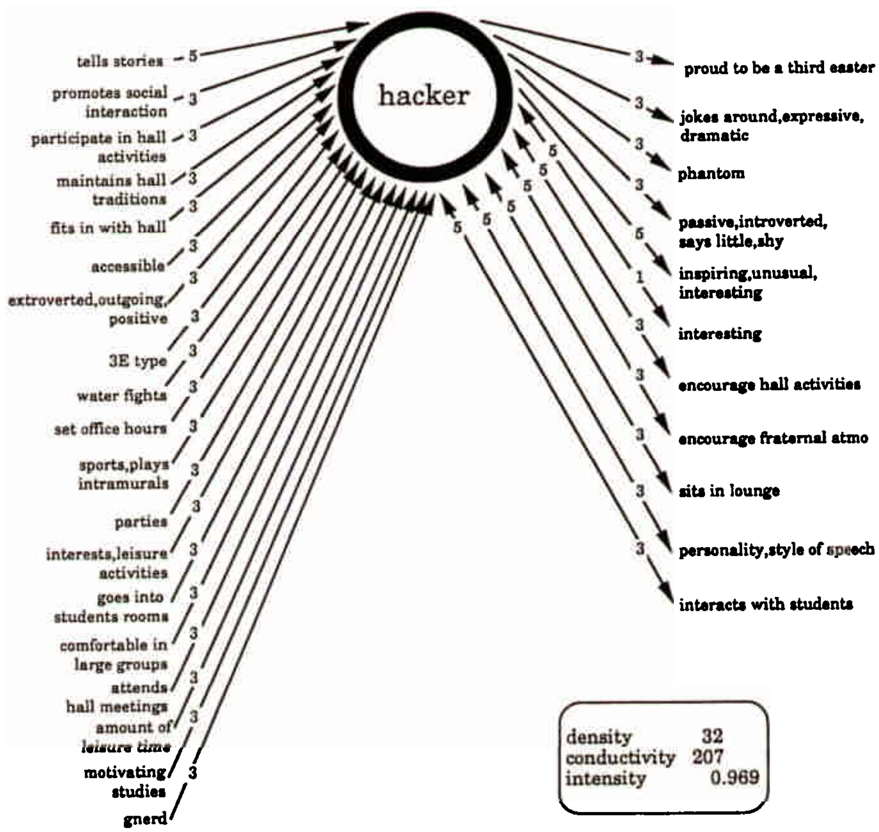

FIG. 4.6. Local network for the concept "hacker."

produced because the extended network for "hacker" is greater than that for "gnerd" for the community at large. In particular, "hacker" has higher extended imageability (18) than does "gnerd" (2). In consequence, presumptions of consensus by students when they label a candidate as a gnerd may be wrong, but similar presumptions based on labeling a candidate as a hacker are more likely to be correct. Stereotypes have great communicative power, but their power in affecting the decision making process may be more a function of their extended position than their local position in the structure of language.

Figure 4.8 shows the conceptual network representing the consensus as to what was wanted in a tutor. No arrows are shown as all relationships are bidirectional. Whether the relationship is positive or negative is indicated by a solid or dashed line. The type for each concept is marked. This map is based on the intersection of all of the students' maps at the end of the tutor selection period. Each concept and all relationships shown were used by every student

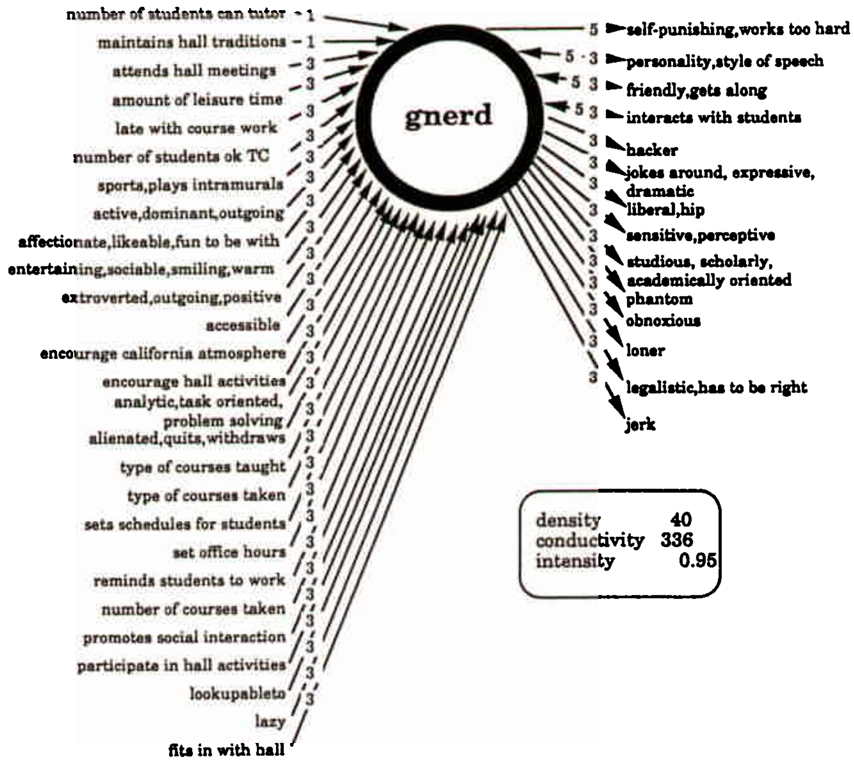

FIG. 4.7. Local network for the concept "gnerd."

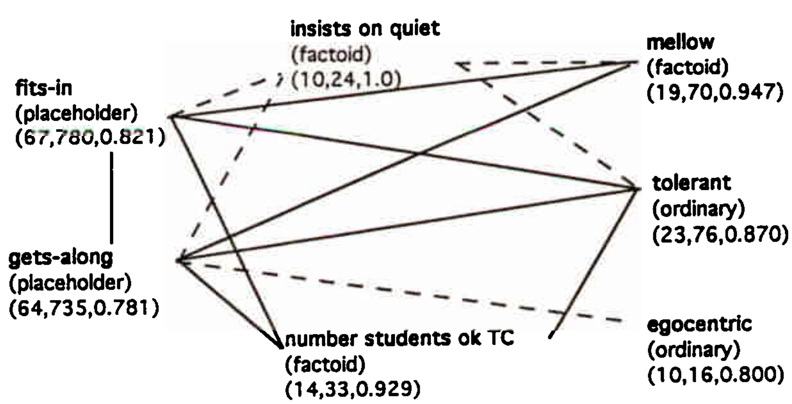

FIG. 4.8. Shared conceptual network for tutor.

in describing what they wanted in a tutor (Carley, 1984, 1986a). Each concept is annotated with information on its position in the typology and the local values for density, conductivity, and intensity. Two of the central concepts, "fits in" and "gets along," are place-holders. Although they have complex meanings, the meaning can not be agreed upon. The students can use these words with assurance that others will agree with them that the candidate should get along or fit in, without agreeing as to what this means. There are, however, three factoids that all students agree help to define what it means to fit in or get along: insisting on quietness, being mellow, and being "ok" from the perspective of many students. These factoids constrain the definition of the place-holders and provide the students with a common set of questions they can use to interrogate candidates, such that all students will agree on the meaning of the answers to those questions.

It is interesting to note that three of the four concepts specified by "gnerd" are major determinants of whether or not someone is chosen as a tutor due to their high levels of evokability. These concepts are "personality," "friendly, gets along," and "interacts with students." These three concepts are specified negatively by the concept "gnerd." Consequently, if a candidate is referred to as a gnerd then he or she is inferred not to have these traits and so does not meet the requirements for being a tutor.

CONCLUSION

A procedure that allows the researcher to begin to relate the nature of language to action has been explored. A model of language was proposed. According to this model, language is a chronicle of social life formed through a lossy-integration process as knowledge is articulated during interactions. Under this model the concept in isolation is meaningless; hence, the smallest unit of meaning is the statement (i.e., two concepts and the relationship between them) or sentence. Representing language as a series of statements makes it possible to discover regularities in concepts' positions in conceptual networks. By using a set of simple dimensions it is possible to depict talk in a society or individual. Such an analysis can point to the communicative power of concepts. Simple analyses show that most concepts are ordinary (i.e., low on all dimensions). Those concepts that stand out on at least one dimension may have particular communicative significance; that is, they may engender consensus or miscommunication. The end result of the analyses is a taxonomy of concept usage, based on structural relationships, that describes the relationship of roles to stereotypes, historical prototypes, and so on.

From a methodological point of view, the argument is straightforward. Texts can be coded as conceptual networks. These networks can, depending on the coding scheme, represent mental models. Coding texts as mental models focuses the researcher on the analysis of meaning. Coding texts as networks allows the researcher to evaluate the texts in terms of the positional properties of the concepts. Examining concept positions focuses the researcher on the communicative power of the concepts. By analyzing language in terms of the positional properties of concepts collected relative to a particular social task, insight might be gained into aspects of the task that have to do with the evolution of social knowledge. The collection of a task-specific language and the subsequent analysis of it using connective and positional properties may also provide insight into the type of concepts used by the subjects to promote and maintain consensus and to communicate effectively.

This research illustrates an approach to studying the relationship between language and society. The methodology, although exploratory, is completely general. Because this work is exploratory, the ideas presented should be tested on other data sets. For example, one might test whether concepts' structural characteristics (i.e., their score on the various properties) are not fixed but dependent on the sociocultural environment and the task being performed.

We see some evidence for this hypothesis in the MIT study. Consider the set of concepts in the MIT data set that are referred to as stereotypes—"hacker," "gnerd," "phantom," and "expert." If all the language for the United States had been amassed, the structural properties of these concepts might change; for example, the concept "phantom" probably would not stand out as a stereotype. Along with this change in structural property comes a corresponding change in meaning. For example, for the United States as a whole, the concept "phantom" might have the dominant meaning "ghost"—that connection might be agreed to by nearly everyone. For Third Easters, in contrast, the dominant meaning is "someone who is never seen out in the hallway."

Another extension from this research would be to determine, for a specific task, whether or not the differences between the structure of an individual's language and the structure of the social language can be used to predict task-related behavior. An additional extension would be to determine whether there is a taxonomy for individuals' language that is similar to the general taxonomy located herein. Another extension would be to explore the evolution of concepts along the dimensions suggested. Palmquist, Carley, and Dale (this volume) show an evolution from the vague to the detailed. In the MIT study (Carley, 1986a, 1986b), there was an evolution from the general to the historically particular. See Kaufer and Carley (1993a) for further discussion of the evolution of meaning.

Language as social chronicle implicitly contains the socially accepted meaning or definitions of the various concepts in the social vocabulary.

Social meaning for a particular concept has been identified with various network measures–richness, imageability, and density. Which particular operationalization is the best is open to debate. However, the proposed model of language may provide a basis for a social theory of meaning. By defining meaning as the definition of a concept or as the network generated when a concept is used, then it follows that meaning is a constructed phenomenon. Another consequence is that the individual's meaning is different from the social or average meaning. All such meanings change over time. Therefore, meaning has a social reality external to any one individual; there may not be a single individual in the unit who ascribes to that definition in total. Coding texts as conceptual networks, and then analyzing these networks using the dimensions proposed herein, helps the researcher to examine the constructed nature of meaning, to determine the basis for individual and social differences in meaning, and to examine the relationship between concept usage and action.