Coding Collected Research Papers to Create a Simple Information System

James J. Sosnoski, Department of Communication, University of Illinois at Chicago, Chicago, IL, USA

Gordon S. Carlson, Department of Communication Studies, Fort Hays State University, Hays, KS, USA

ABSTRACT

A major issue with research archives is the difficulty of finding contextually relevant information when developing a theory or project. Utilizing a sociotechnological approach, this paper proposes a descriptive markup language to function as a notational system to be used to convert collections of research papers in various disciplines into an information system. Markup codes attached to research papers can significantly improve data retrieval. A theoretical framework for the notational system is proposed, drawn from Charles Fillmore's frame semantics and Thomas Kuhn's conception of a lexicon specific to numerous research areas. Examples of how the notational system can be used are offered and the steps needed for the implementation of the notational system are outlined.

Keywords: *Descriptive Markup Language, Information System, Markup Codes, Notational System, Research Papers*

INTRODUCTION

Digital libraries of research publications are usually spread across various online sites. And these sites are not structured in the same way. This makes the retrieval of data that goes beyond keyword searches from them difficult—e.g., relations among concepts forming a model of the situation, conceptual changes in concepts drawn from previous research.

To make these issues more concrete, consider the digital libraries that contain publications about communication research. Because it does not have a uniform naming pattern, communication research is enigmatic. It is, for instance, difficult to relate the assumptions underlying research projects to each other. Consulting bibliographies is often futile. The titles of research publications in communication studies are not related semantically. For

DOI: 10.4018/ijskd.2013100104

example, agenda setting theory is not semantically related to framing theory which is often considered a species of agenda setting theory. Looking only at the terms, the conceptions would not seem connected. On the other hand, speech codes theory would seem to be related to speech act theory because they are semantically related. However, Philipsen's (1997) conception of "codes" is quite different than Searle's (1969) conception of "acts." Similarly, Kenneth Burke's concept of symbolic action (1966) is semantically related to and considered an instance of symbolic interactionism. However, Burke views communication from the point of view of its author in one-way transactions and Mead/Blumer (1969) emphasize the dialogical character of communication. There is no semantic connection between standpoint theory and action assembly theory, yet conceptually they are more closely related than speech acts and speech codes. There are many other instances of the absence of cohesiveness in the discourses of communication researchers that can illustrate the difficulty of finding patterns of interrelations in their conceptions. Semantic coherence, however, is not the problem. The problem is the lack of a patterned, descriptive system capable of marking the similarities and differences among the theories in the research archives more precisely.

Why is this a problem and how serious is it? Whenever publications reach a certain mass, they go beyond the capability of humans to remember where they are so that they can be found when needed. Take a simple example: a person begins to purchase books. At first, she has no difficulty in locating a book which she needs. With even a small bookshelf, at some point she will have to put her books in order to form a library. To do so, she must categorize them. Let's say she organizes her books in the categories: very interesting, mildly interesting, not very interesting, and not at all interesting. This idiosyncratic system might work well for her in as much as she would be likely to refer more often to the very interesting books than the totally uninteresting books.

The problem with an idiosyncratic filing system is that it is suited to an individual. The categories are meaningful to that individual but are not shared by specific groups. Categorization is often at the heart of data retrieval. Since it would not be feasible to change the names of communication theories, we believe that there is a need for a notational system for describing communication research that could be used as a markup language. Such a notational system would not be in any way prescriptive. It would be designed to show the inter-relatedness of communication research.

We can't predict what the mark up system will provide; we can only create the framework and then through using it see what others will find. Put another way, we seek to avoid a framework of notation which is deterministic with predictable outcomes because this will lead to a narrower set of possible outcomes. With less potential outcomes it is possible for "false positives" where two documents which are not the same are marked up in such a similar way that a user would be confused or misled. While this problem of naming conceptions exists in all research fields, we use the field of Communication Studies (referred to as communication) as a case study since it draws from so many other fields and is a nexus for many competing and contradictory concepts. The problem of the idiosyncratic naming of communication research and our proposed remedy are sociotechnological and directly impact the creation and management of knowledge. Solutions to this problem will lie at the intersection between people and technology as a once necessary print culture fades behind new technological information management regimes that allow people to interact with information in radically divergent ways.

Data retrieval is a common problem in large organizations. In many instances, the data and its organizing technologies are structured in ways that make it difficult to locate the information you need; for example, papers about contextually similar situations. From our point of view, the problem is in the way the data in a digital

library of publications, memos, or letters is coded. Our intention in this paper is to describe a way of designing a notational system that can be applied to collections of papers so that their contents can be searched for the purpose of comparing models of the pertinent conceptual domain (the model and its components). To illustrate the notational system, we describe a way of coding a digital library of papers on communication theory, which is the conceptual domain in which we work.

This problem came to our attention while developing a learning environment that emphasizes context of use and conceptual change in research about human communication. Although the code we propose is tailored to searching publications in this conceptual domain, we believe that this type of coding can also be applied to conceptual domains pertinent to other organizations whether social, commercial, or educational. Most research publications in the social sciences and in business concern recognizable "situations" common to experiences of everyday environments.

Problematic situations that occur in organizational communication — for example, sexual harassment — have common elements: persons, settings, language use, purposes, and effects. These elements are common to all experiential situations whether in a business organization or in an educational institution. Variations of the relations of these elements to each other characterize types of situations. This has been demonstrated in Charles Fillmore's work on frame semantics. Since research publications are discourses, these elements are discernible and can be used as the basis of a notational system in a variety of conceptual domains.

By conceptual domain we mean a body of information about a general conception in which the various embedded concepts can be related to each other to form a mental model of it. For example, consider the word "computers" as a conceptual domain. Computers, of course, have various parts that are related to each other through a generic model of a computer that can be envisioned and described. If you owned a computer manufacturing company, there no doubt would be a database available to you with various fields for the sales side of the business. Chances are, however, that although the company may own a digital library about computers, it probably would not be searchable in the same way an academic text database can be searched due to the significantly different contexts in which each is deployed.

We conceptualize the issue of connecting research discourses in communication research as a sociotechnological system that must balance the increasing ability of computational systems and "big data" collections against the social and human interests in scholarly research. We agree with Enid Mumford (1989, 1996, 2003) and her peers that the interaction of social and technical systems must be optimized and do so in a way that benefits research goals rather than technological ones.

To serve this purpose, a notational system could be based on the context of use and the semantic relations of the conceptual models and their components that have been used by researchers. Since all communication occurs in a context, any research into communication events is related to its context of use, that is, to the situation it addresses. In Charles Fillmore's frame semantics linguistic expressions are either explicitly or implicitly related to situations (2006). This factor can be exploited in designing a markup code for published papers in digital form.

In Fillmore's view of "frame semantics," sentences (including statements) invariably imply situations and, as mentioned earlier, the elements common to them. An example of this phenomenon is a "business transaction," a situation with the common elements of "buyers," "sellers," "goods," etc.:

In particular, I tried to show that a large and important set of English verbs could be seen as semantically related to each other by virtue of the different ways in which they 'indexed' or 'evoked' the same general 'scene'. The elements of this schematic scene included a person

interested in exchanging money for goods (the Buyer), a person interested in exchanging goods for money (the Seller), the goods which the Buyer did or could acquire (the Goods), and the money acquired (or sought) by the seller (the Money). Using the terms of this framework, it was then possible to say that the verb buy focuses on the actions of the Buyer with respect to the Goods, backgrounding the Seller and the Money; that the verb sell focuses on the actions of the Seller with respect to the Goods, backgrounding the Buyer and the Money; that the verb pay focuses on the actions of the Buyer with respect to both the Money and the Seller, backgrounding the Goods, and so on, with such verbs as spend, cost, charge, and a number of others somewhat more peripheral to these. Again, the point of the description was to argue that nobody could be said to know the meanings of these verbs who did not know the details of the kind of scene which provided the background and motivation for the categories which these words represent. Using the word 'frame" for the structured way in which the scene is presented or remembered, we can say that the frame structures the word-meanings, and that the word 'evokes' the frame. ("Frame Semantics," 378)

Fillmore's example includes fields that would normally be used in a relational database of a business. As we hope to show in what follows, the situation in which a communicating event occurs, explicitly or implicitly, similarly evokes relations among its components and thus can be used as the basis of a markup language.

PARADIGMS AND LEXICONS

We have grown accustomed to thinking of research publications as applications of theoretical models which suggests that the concepts that constitute them belong to a conceptual system and cannot be understood apart from it. In some respects, Kuhn's conception of paradigms in his *The Structure of Scientific Revolutions* (1965) contributed to the view of theory as a theoretical system in which the relation between the concepts that comprised it were fixed until a paradigm shift was given status by its publication in important journals or by important presses.

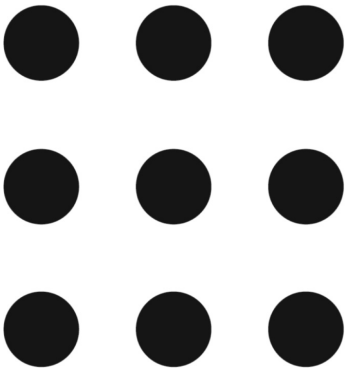
Kuhn's later view of scientific theories as lexicons is largely neglected in favor of the paradigm view to which considerable attention had been paid. The paradigm view places emphasis on theories as systems in which the concepts that comprise them must be understood in the context of the theories. The lexicon view which he developed in his writings after 1990 detaches the concepts from the theories and construes them in terms of the network of interrelated choices of words available for use that a lexicon presents. In *The Road Since Structure*, his presidential address to the Philosophy of Science Association in October 1990, Kuhn argued:

Given a lexical taxonomy, or what I'll mostly now call simply a lexicon, there are all sorts of different statements that can be made, and all sorts of theories that can be developed. Standard techniques will lead to some of these being accepted as true, others rejected as false. But there are also statements which could be made, theories which could be developed, within some other taxonomy but which cannot be made with this one, and vice versa. (T. S. Kuhn, et al., 2000 published posthumously)

The shift from thinking about research as theoretical systems to thinking about the concepts used in research as a lexicon can be considered an instance of "thinking outside the box." To illustrate such a shift, we turn to a famous instance of the need to think outside the box, namely, the 9 dot puzzle (Figure 1).

Figure 1 shows nine dots arranged in a set of three rows. Your challenge is to draw four straight lines which go through the middle of all of the dots without taking the pencil off the

Figure 1. The 9 dot puzzle



paper. If you are using a pencil, you must start from any position and draw the lines one after the other without taking your pencil off the page. Each line starts where the last line finishes.

A popular riddle addressing the cognitive structures people use in solving problems, the 9 Dot Puzzle provides a clue to understanding the shift from paradigms to lexicons. As in the puzzle, the problem is how to relate the dots (or the theories) to each other in a rule governed (or systematic) way. While the 9 Dot Puzzle illustrates the problem of finding a pattern that relates the theoretical components of a conceptual domain to each other, it also suggests a solution — "thinking outside the box."

Keeping in mind that categorization is often central to data retrieval methods, we need first to consider the boxes (categories) into which research is placed. Continuing with our case study of communication, we conducted an examination of the organizational schemes of five popular textbooks in communication theory revealing the lack of systematic uniformity in categorizing communication theories:

- *A First Look At Communication Theory* (2012, 8th edition) authored by Em Griffin;
- *Theories of Human Communication* (2011, 10th edition) authored by Stephen W. Littlejohn and Karen A. Foss;
- *Introducing Communication Theory: Analysis and Application* (2010, 4th edition) authored by Richard West and Lynn H. Turner;
- *Communication Theories: Perspectives, Processes, and Contexts* (2005, 2nd edition) authored by Katherine Miller;
- *Building Communication Theory* (2003, 4th edition) authored by Dominic A Infante, Andrews S. Rancer, and Deanna F. Womack.

Table 1 illustrates major communication theories or concepts discussed in the five textbooks. Columns represent each text and rows represent the various concepts as labeled and explicated in each text. For example, Miller is the only text to directly address Symbolic Organization but three texts (in row 3) discuss comparable conceptions of messages but different labels. The table demonstrates the lack of consistency in naming theories or concepts in communication research.

The chart of the various categories the authors of communication theory textbooks employ shows some family resemblances but, even at the surface, semantic differences appear significant. For example, is the communicator "the self" and is "trait theory" (Littlejohn/Foss) to be understood as a "traits approach"? Or,

Table 1. Chart of the organization of communication theories in popular textbooks

Griffin	Littlejohn & Foss	West & Turner	Miller	Infante et al.
	Symbolic Organization			
				Traits Approach
	The Communicator & The Message	The Self & Messages	Message Production & Message Processing	
Interpersonal Communication		Interpersonal Contexts		
	The Conversation			
	Persuasion Approach			
	Discourse and Interaction			
	Verbal Behavior			
	Nonverbal Behavior			
Influence				
Relationship Development & Relationship Maintenance	The Relationship	Relationship Development	Communication in Developing Relationships & Communication in Ongoing Relationships	
Group Communication	The Group	Groups & Organizations	Small Group Communication	Group Contexts
Organizational Communication	The Organization		Organizational Communication	Organizational Contexts
Public Rhetoric		The Public		
Gender & Communication				
Media & Culture	The Media	The Media	Media Processing & Effects; Media and Society	Mass Media Contexts
	Culture & Diversity	Culture & Communication		
Intercultural communication				

is "media" (Littlejohn/Foss & West/Turner) equivalent to "media processing & effects" (Miller) as well as to "mass media contexts" (Infante, et al.)? Perhaps most interesting, why are there so many blanks in the chart?

A careful examination of the available textbooks in communication theory would reveal that many theories do not appear in the same categories in them and many theories that appear in one textbook do not appear in the others.

THINKING ABOUT COMMUNICATION THEORY OUTSIDE OF THESE CATEGORICAL BOXES

As a thought experiment, let us consider what would happen if we disassociated the theories from the authors who developed them, then stripped away the traditions attributed to them, and finally, disconnected the concepts embedded in these theories. In other words, what if we took communication theories out the theoretical boxes (categories) into which they are customarily put? The result of this thought experiment would be an enormous mass of conceptions cut loose from their originating purposes in research projects. It would be something close to what Thomas Kuhn, drawing upon John Lyons' *Semantics* (1977), called a "lexical taxonomy," a view that replaced his earlier description of the structures of scientific inquiry as paradigms (2000). We would be left, returning to our opening analogy, with a giant "dot" puzzle.

What could justify this dismantling of theories? There are several important reasons:

1. Theoretical models are not "field-wide" and pertain only to very particular situations in which problems are encountered. They have to be changed to pertain to other situations meaning that their embedded concepts have to be changed, new ones added to the mix, and old ones dropped out altogether. Strong theoretical models are often contextual;
2. Theoretical models are systems to the extent that the inter-relations of the concepts they contain are bound together with respect to the use of the model which is *tied* to the situations they address. Theories do not "travel." When their concepts are substantially re-conceptualized to fit dissimilar situations, technically, they need to be understood as different theories even when the concepts, terms, or names are retained. Similarly, when the relations among concepts change, the theories they constitute change; that is, they are *new* theories even if they are referred to by the same name as the ones they replaced;
3. The conceptions in models are not "owned" by the authors of the theories involved and can be made available to researchers who are in the process of building theoretical models as heuristic guides for their projects.

THOMAS KUHN ON SCIENTIFIC THEORIES AS CONCEPTUAL PUZZLES IN LANGUAGE GAMES

Thomas Kuhn retained his conception of scientific inquiry as puzzle-solving throughout his career (*The Structure of Scientific Revolutions*, ch.5, 1970, 35-42). Though he abandoned his conception of a paradigm and changed his conception of incommensurability after 1990, he retained the idea of science as puzzle solving. In *The Road since Structure*, he proclaimed: "what scientists do is solve puzzles" (1990, 96). They do this by formulating theories, which we have already noted, are made up of conceptions.

In this context, Kuhn's historically oriented philosophy of science is meta-puzzle solving in the sense that he tried to solve the puzzle of conceptual change in scientific inquiry in general. He wanted to find a pattern in the development of scientific theories. Scientific theories —communication theories among them — are larger units than conceptions and provide a framework for them. If we broke up theories into conceptions, as suggested earlier, then the conceptions would no longer be contained by frameworks and would, so to speak, be taken out of their "boxes." If conceptions were removed from their constraining theories, we would be able to treat them as conceptions of specific relations among the components of communicating.

This is roughly what Kuhn does in his "linguistic turn" in which he regards theoretical concepts as categories in a lexicon that structures the conceptions it includes:

Conceive the lexicon as a module within the head of an individual group member. It can then be shown ... that what characterizes members of the group is possession not of identical lexicons, but of mutually congruent ones, of lexicons with the same structure. ***The lexical structure which characterizes a group is more abstract than, different in kind from, the individual lexicons or mental modules which embody it. And it is only that structure, not its various individual embodiments, that members of the community must share.*** (emphasis ours, 2000, p.104)

Historically, scientists develop conceptions and relate them to each other in theories. Taken together they form various lexicons roughly equivalent to the conceptual domains that are usually marked off from each other as disciplines. This is evidenced by the circumstance that each discipline, or perhaps more accurately sub-discipline, has at least one terminology (read lexicon) which functions as a language that its members use to collaborate and which is often unintelligible to members of other disciplines.

In *The Road Since Structure*, Kuhn argues that the history of science is a history of conceptual changes (91); that normal science is a stage in which the structure of the lexicon being used is not substantially changed; and that scientific revolutions occur when the lexicon substantially changes its structure—the conceptions in it need to be re-understood or replaced even when the concept-terms are retained. From this perspective scientists solve puzzles by changing the relations among the conceptions they use. Some changes affect the entire lexicon, others do not. Sosnoski and Carlson (2013) have coined the term Conceptual Logistics for this process of connecting and re-conceptualizing concepts in research discourses.

Consider the dots in the 9 Dot Puzzle in Figure 2 to symbolize conceptions and the arrangement of their three rows to symbolize the theoretical model that includes them. Further, consider the lines that connect them to symbolize the "structure" of the theory since it shows how the dots, symbolizing concepts, are connected.

Consider connecting the dots "outside of the box" (see Figure 3).

If we continue to construe the dots to constitute the components of a theoretical model and the lines that connect them to symbolize their structure, when we retain the rule that requires the person trying to solve the puzzle *not* lift the pencil from the paper, the structure of the "theory" does not change even though its configuration does.

The different configurations of the 9 Dot Puzzle in Figure 4 do not substantially change the structure of the lines which look somewhat like a kite in the wind. The lines start

Figure 2. Connecting the dots "inside the box"

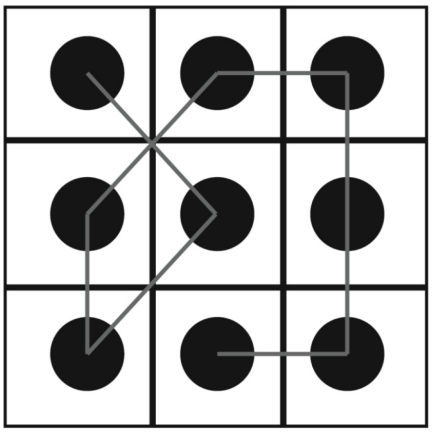


Figure 3. Connecting the dots "outside the box"

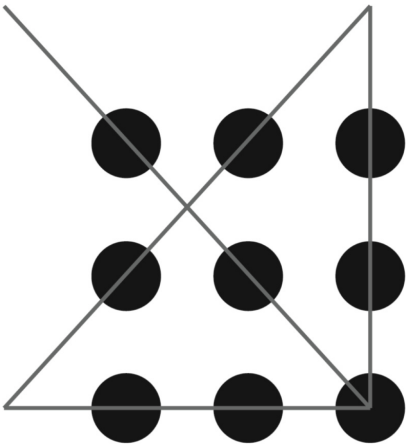
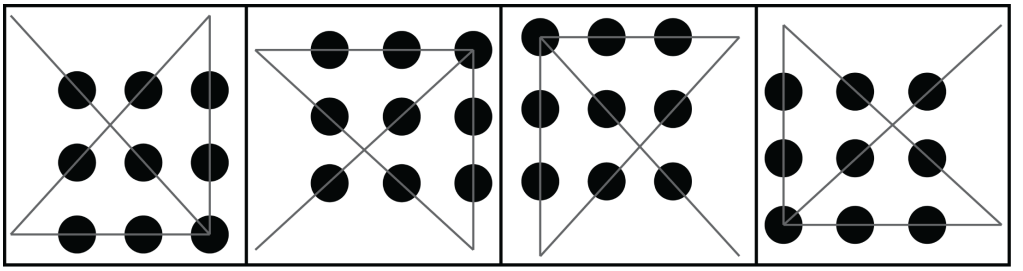


Figure 4. Solutions to the 9 dot puzzle



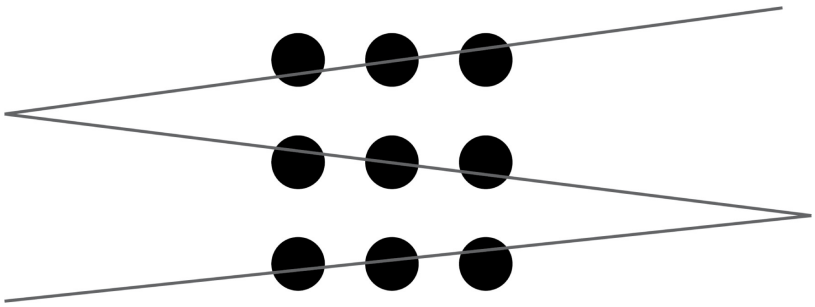
from different corners but make the same sort of connections through the center of the dot without lifting the pencil. Changing the rule substantially results in quite different structures as seen in Figure 5.

Though this analogy is rather crude, it symbolizes the difference between normal and revolutionary scientific inquiry in Kuhn's view. Considering language as a game, as Wittgenstein proposed (1953) and Kuhn accepts, the configurations in Figure 5 require a change in the rules of the game to be "valid."

Now we have two ways of operating outside the box: on the one hand, we can disassociate the dots from the constraining category of a box, and, on the other, we can also change the rules of the game. By working outside the framework of a theory, we can explore other possible ways to organize the conceptions detached from their theory "boxes." In addition, we can change the rules of the game. In our case, the rule change is from a view of language as independent of experience to one in which language depends on experience:

The main feature that distinguishes Cognitive Linguistics from generative grammar has to do with the place of meaning in the theory. In the generative model the structure of linguistic expressions is deemed to be determined by a formal rule system that is largely independent of meaning. By contrast cognitivists argue that

Figure 5. Alternative solution to the 9 dot puzzle



linguistic structure is a direct reflex of cognition in the sense that a particular linguistic expression is associated with a particular way of conceptualising a given situation. This leads to a quite different view of the relationship between language and cognition in general. Whereas generative grammarians claim that there exists a rich set of principles of language design (Universal Grammar) that are specific to language, the cognitivists believe that, although universal principles governing the design of all languages may well exist, they will eventually be found to be rooted in cognition. (Lee, 2001, 1)

This "rule change" allows us to understand the use of language as *always* situated (Brown & Yule, 1987; Gee, 1999; Georgakopoulou, 2004; Fillmore, 2006; Halliday, 2009). Given *this* uniform pattern, it is possible to code a theoretical lexicon made up of conceptions no longer attached to theories because, from a cognitive linguistic standpoint, they retain the meaningful ways they conceptualize a given situation. From this perspective, the theoretical concepts used in research projects are always conceptions used by someone to conceptualize something in some place at some time, more often than not, collaboratively. In the next section, Charles Fillmore's frame semantics provides a guide to describing conceptions with a notational system based on a constant component of a communicating event: the situation.

A FUNCTIONAL MODEL OF A SITUATED COMMUNICATION EVENT

Ordinarily, models of communication are based on the *sine qua non* components of communicating. Typically, five components are regarded as necessary for a communicating event to occur. From an historical point of view, they were initially identified by Aristotle in his explanation of change as the result of four "causes": efficient, formal, material, and final (Charlton, 1970). Subsequently, a fifth cause was added — instrumental. Using traditional terms, we stipulate that the following five components of communicating can be found in most models of communication.

This model is often said to have its origins in the Shannon/Weaver (1949) model which is linear. The "traditional" model (Table 2) is usually associated with Wilbur Schramm (1971, 1954) who modified the Shannon/Weaver model by adding a dialogical aspect (represented by the "sender/receiver" component suggesting changes in roles by the persons involved). It also adds "channel" as a component because channels alter messages.

For the purpose of developing a markup language, in contrast to the traditional model, we construe communicating as a situational system in the sense that each component *functions* in relation to the other components to produce a

Table 2. Traditional components of communication models

Code		
sender/receiver	message/channel/noise	receiver/sender
Context		

communicating event. We re-conceptualize the process of communication from a functional point of view (Halliday & Kress, 1976), and rename the conceptual components to reflect their *function* in the system (see Table 3).

Since persons are involved, situations are involved since persons do not exist in a vacuum.

In a communicating situation, the persons involved invariably bring to it their cognitive frameworks. Two are especially critical, namely, the functions of the semantic and episodic memory systems (Reisberg, 2010; Tulving & Craik, 2000). Every person involved in a communicating event is influenced by his or her memory systems. The semantic memory stores concepts and the episodic memory stores experiences. Both memory systems are private and not accessible to observers. However, the sign systems used in communicating reflect them. If a communicator uses a particular concept or refers to a particular experience, we do not know precisely how she understands or experiences the events. We only have access to the public signs produced. Unlike the traditional model which identifies only the public components of communicating, the functional system we are using includes four "private" functions (Table 4) to account for the cognitive aspects of communicating events. Including the private spheres adds four functions to the model.

Table 4 describes a moment in a communicating event. We use the expression "a *moment* of a communicating *event*" to allow for the circumstance that the persons in the event often change roles in the process of communicating. Our model represents the equivalent of a turn in a conversation where person A is addressing person B. In the next moment, it is B's turn to address A. The same model applies to the second turn. To develop the notational system we are proposing does not require a more complex model.

In any *moment* of a communicating event the persons who are involved bring to the public event different cognitive frameworks which are private. The cognitive frameworks of each

Table 3. A functional view of communication events

	cultural codes	
communicator >	text	< communicate
	reference to experience (situation)	

Table 4. A functional cognitive model of communication events

Public Elements: Communicator *Private Elements:* Communicator's Pretext Communicator's Context	Cultural Codes Text Reference to Experience (Situation)	*Public Elements:* Communicatee *Private Elements:* Communicatee's Pretext Communicatee's Context

person differ from all the other person(s) in the interaction. Thus, they need to be considered separate functions in a communicating event — the communicator's pretext and context are distinct from the communicatee's pretext and context in the dyadic model we've drawn. Since communication theory often involves conceptions that attribute meanings to those involved in the event by using terms such as "intends," "means," "feels," and so on, it is necessary to consider the cognitive functions of the private components of communicating before we can dissect the system. Their functions are to contextualize (to locate an experience in epistemic memory) and to pre-textualize (to locate a concept in semantic memory) textual data. Both concepts and experiences are activated by a text but they activate concepts and experiences that existed in the communicator's or communicatee's minds *before* the text is constructed or reconstructed.

In addition, the conception of a text in our functional model approximates the conception of a channel of communication in the traditional model. We borrow this view from cognitive linguistics where the text is in itself meaningless apart from a communicative interaction (Langacker, 2002, p. 5). Further it is multimodal (as are channels) and can be designed as a verbal, visual, aural artifact or any combination of the sense activating aspects of a channel. A text is no more than transmitted public symbols that, as textual cues, activate semantic and contextual memory systems in communicatees (and also in the researchers studying the interaction).

The model identified in Table 4 is the premise for our systematic approach to describing communication events from a sociotechnical perspective through a notation system. With this model of communication events we have the ability to more systematically construct new communication theories and more usefully describe and analyze ones current in the research discourse. To accomplish these goals, comparisons are required. New theories are developed by comparison with older ones and analyzing theories cannot be done in a theoretical vacuum. Given these goals, it may seem counter-intuitive to dismantle available theories, cutting loose embedded concepts from their conceptual frameworks; however, theories are rarely built *in toto* from scratch. They are usually developed by assembling concepts into interrelated conceptions. In the same vein, analyzing a theory, by definition, means breaking it up into its parts.

Before we explain the theoretical construction of our functional model as a template for our notational system, we need to consider the changes that the various conceptions of communication undergo in the course of scientific inquiry and the relationship of these changes to a lexicon.

CONCEPTUAL CHANGES

Since conceptions are always situated (always referring to some situation either explicitly or implicitly), situation, a constant component of a communicating event, is a common denominator and can be used as the basis of a code or notational system. But, as Kuhn argues in *Structure*, conceptual change is often a re-conception of a situation. Also, in his post-*Structure* view of the history of science as an evolving lexicon, conceptions are understood to change over time. Some of their elements disappear and are replaced by "newer" ones, altering the elements that are retained. Competing theories, composed of differing individual concepts and structures, also permeate the research discourse about human communication. The question then arises: if the conceptions of situations change, how does this affect our functional model? It doesn't. Kuhn does not focus on conceptual changes that occur *because the situations change*. Nonetheless, conceptual change occurs in both instances and conceptions are situated in both. As he notes in his presidential address:

The lexical structure which characterizes a group is more abstract than, different in kind from, the individual lexicons or mental modules

which embody it. And it is only that structure, not its various individual embodiments, that members of the community must share. (2000, 104)

Our model is "more abstract" than an individual conception of it. Conceptual Logistics (Sosnoski & Carlson, 2013), which is a theory of conceptual change, provides a framework for connecting and understanding these theories across diverse contexts of use. Outside of the researchers who study conceptual change, it does not factor into "normal" scientific inquiry. Not surprisingly, a practical approach to systematically analyzing this aspect of research discourse simply does not exist, in part because the process of constructing and disseminating research using print culture tools cannot easily accommodate it.

We advocate a sociotechnological approach employing a notational system as a markup language to provide the basis for a down-to-earth information system that can bridge the social and technical gap within organizations concerned with knowledge development, for instance, the American-based National and International Communication Associations. In order to make the case for our proposed notation system we provided a functional model of communication that operates as a framework ripe for systematic investigation and the type of markup languages that computers and people alike can use in understanding and accessing knowledges derived from communication research.

With these considerations in mind, we propose a notation system for creating a markup language that can be applied to research discourses about human communication.

THE NOTATIONAL SYSTEM

Using the model we outlined in Table 4 we can begin constructing the codes of a notational system. In the Table 5, we introduce abbreviated forms of the nine constant functions which are also the core of the proposed code. The code abbreviates the names of the functions and also the factors that modify them in particular situations. The advantage of using abbreviations is that, once they are learned, they bring to mind the terms to which they correspond. In order to keep the codes short, the core functions are reduced to a single letter when they modify other core functions. For example, the function of communicator is abbreviated to "tor" and of communicatee to "tee," but in order to distinguish the pretexts and contexts of communicators from those of communicatees, the letter "r" abbreviates "tor" and the letter "e" abbreviates "tee" resulting in the expressions "r-pxt," "r-cxt," "e-pxt" and "e-cxt." Similarly, the letters "C" and "S" abbreviate cultural codes and situations as linguistic elements in the text (which is abbreviated to "txt"):

- **The Communicator Function:** The function of communicators is to communicate. Whatever they communicate is influenced by the conceptual frameworks existing in their semantic memories which are not directly accessible but can be accessed by inferences about the significations made public in the texts persons construct. Similarly, the context of persons' experiences are stored in their episodic memories and are not directly accessible except by inferences

Table 5. The nine fundamental codes

r-pxt = Communicator's Pretext	C} = Cultural Codes	e-pxt = Communicatee's Pretext
tor = Communicator	txt = Text	tee = < Communicatee
r-cxt = Communicator's Context	S} = Ref to Situation	e-cxt = Communicatee's Context

about recognizable situations made public in the texts persons construct.

Given the circumstance that neither the communicator's nor the communicatee's pretexts and contexts are available directly, their functions in a communication event have to be described with the terms used for public cultural codes and references to recognizable situations the coder identifies in the text as an interpreter.

The main function of the communicator is to construct a text. "C}" indicates that the communicator was affected emotionally or cognitively and "S}" indicates that the references to experiential situations provide a contextual framework for understanding the text. The impact of pretexts and contexts on communicators is to dispose them to construct the text in a certain way. Though these are constant components of communicating, they are not always included in the theoretical models used in research:

- **The Text Function:** As we noted earlier, texts are carriers of meaning and not meaningful in themselves. A text in a communicating event is always "activated" by a communicator or a communicatee. An important aspect of texts is their modality. Texts are sensible phenomenon and convey meaning mostly through sounds or images though in some situations smells, touches, and tastes contribute to the communicating event. Textual modes modify the meaning of a communication which may be further modified by the media involved. "@" indicates that the communicator's text and the communicatee's text are altered when modified by their media or modes. Texts are characterized by the modalities in which they are constructed, or, in the case of communicatees, re-constructed. The typical types of modifications are seen in Table 6.

Generally, the impacts of modality will be coded:

- Verbal (words) = w@txt;
- Visual = v@txt;
- Musical = m@txt;
- Gestural = g@txt;
- Gustatory = t@txt.

Multimodal texts will be coded according to the combination of modalities in their construction, for example:

- verbal & visual = wv@text;
- verbal, visual, & musical = wvm@txt.

In instances where the modalities depend on the circumstances in which they are constructed or re-constructed, the coding would be: wvm@text "&" v@txt, etc. — for example if the communicatee reconstructed a multimodal PowerPoint presentation from a printout of it:

Table 6. Modality modifications

Sensing	Modalities	Cognitive Abilities	Media
Seeing	Visual	Reading Viewing	Words Images, Gestures
Hearing	Auditory	Listening Speaking	Language, Music Language
Touching	Tactile	Distinguishing Touches As Significations	Gestures
Smelling	Olfactory	Distinguishing Smells As Significations	
Tasting	Gustatory	Distinguishing Tastes As Significations	

- **The Communicatee Function:** The function of a communicatee in a communicating event is to interpret the text received, to render it meaningful. This cognitive activity is influenced by the way in which the text affects the communicatee. At any and every moment in the communicating process, communicatees are affected by the text both emotionally and cognitively. As in the case of the communicator(s), the communicatee's pretexts and contexts are not always included in the models of communicating used in research;
- **The "Cultural Code" Function:** Cultural codes are construed here as semantic frameworks that characterize particular discourse communities. The semantic framework is stored in the semantic memory but is a private cognitive activity that is accessible only in the symbols of the text and their signification. The default coding is that the persons in the communicating event share the same *generic* semantic framework. In instances of intercultural communications or other communications in which the cultural codes differ, the communicators and communicatees are modified by C} and {C respectively to indicate that cultural codes are significant factors in the communication according to the researcher. Otherwise, they would not be explicitly coded (this does not violate typical markup language prohibitions on ambiguity and instead represents a default meaning when a specific code is not used);
- **The "Recognizable Situation" Function:** Recognizable situations are "cognitive" events in which past experiences, stored in the epistemic memory as generic or typical, are evoked by present experiences. They correspond to the semantic frameworks but are not identical to it. Like cultural codes, they are cognitive activities that are only accessible in the discursive "textworld" (Werth, 1999) drawn by references in the text to typical experiences. The default coding is that the persons in the communicating event share the same *generic* experiential frameworks. In instances of intercultural communications or other communications in which the experiential frameworks differ, the communicators and communicatees are modified by S} and {S respectively to indicate that "textworlds" are included in conceptions because they are significant factors in the communication according to the researcher. Otherwise, they would not be coded;
- **The "Outcome" Function:** The core functions represent a *moment* in a communication *event*. The expression "a *moment* in a communicating *event*" allows for the circumstance that the persons in the event often change roles in the *process* of communicating. This process is not cyclical. It is not a repetition of the same moment over and over. Rather, the moments in communication events are *sequential* because communicating progresses (in the sense of processes toward a goal). To reflect this, an additional component, effect, needs to be added which usually, but not always, results in a continuation of the communication. Every text presented to another person has some effect on that person which is perceptible as an action (reaction, interaction, disposition to act).

The "effect" or "outcome" component of communicating is not one of the categories so far discussed. This function is not present in a moment of a communication event and occurs between moments of the communicating process and is therefore "separate" (at an interval) from the moments of the process. The "core" communicating event-moment in this model needs to be understood as a module that can be repeated after an interval. In each iteration, the components multiply creating numerous constructs, pre-texts, contexts in the participants via textualization. As communicating events increase in number and expand over time, they have both a sequential and a cumulative dimension. Considering one "turn" in a conversation to be a moment in a communicating event, as the turns are taken, new textual choices, structures,

content, and text worlds are produced. As each new turn takes place, the previous turns frame the participants understanding of the new turn.

The outcome function needs to be included in any description of communicating. It is the result of the various ways in which the communicatee is affected emotionally and cognitively at any given moment in the process. All of the core functions (communicator, communicatee, pretexts, contexts, texts, and effects) are related to other core functions in a communicating event to some extent. There are numerous relations that are described in communication research and need to be coded in descriptions of communicating events.

THE RELATIONSHIP BETWEEN OR AMONG COMPONENTS

The codes symbolizing the relations between and among the nine components are not abbreviated in the same way as earlier indicated. Instead, symbols that suggest relations are employed much as in mathematical equations. However, since we use only the symbols available on a standard computer keyboard, the relationship codes are not as easily translated into the concepts they symbolize. However, the patterns we use are as close to the use of these symbols elsewhere as we can make them. For example, direction is coded by the symbols ">" and "<". An effect or outcome is coded by the symbols "->". When the direction between the communicator and communicatee is interactive, the symbols are "><". When the communicator and communicatee are antagonistic or disagree, the code is "÷" or "<÷>" and the direction is away from each other: "<" communicator directed away from communicatee, and ">" vice versa. Modification of components are symbolized with the codes: "}" and "@." Brackets "[ ]" and "{ }" indicate that the components and their relations are considered as parts of a whole or a system. A symbol in bold font indicates that the component or relation is emphasized in the theoretical model being described (see Box 1 for relationship codes).

Even though the codes are basically abbreviations of concepts or common symbols, they would be obtuse to most persons; however, a simple translation program would convert them into more recognizable expressions and leverage those translation abilities across various technological implementations such as database or "big data" mining systems. This would also be very helpful to non-English speaking persons.

SAMPLE MARKUP

The value of a markup system is in the emergent properties resulting from the analysis the code allows for. Given the number of papers, chapters, books, and articles already in research databases compounded by the rapid increase in digitizing new research, it is likely impossible to predict the number and type of possible results or relationships this coding approach could yield. However, certain findings are more likely than others and here we present some basic examples of our proposed coding system in action. These examples are illustrative and do not represent the extent of possible findings which are, in actuality, an emergent property of the approach.

The examples shown in Box 2 illustrate how the proposed markup language can identify similar conceptions of communicating and enumerate some of the dissimilarities between conceptions.

This coding identifies dissimilarities between agenda setting and face negotiation as communicative activities. Whereas agenda setting forces the communicatees to entertain a specific cognitive framework, face negotiation involves a mutual modification of cognitive frameworks with respect to cultural codes.

CODERS AND FEASIBILITY

It is not feasible to expect organizations to code all the papers in their archives. For example, EBSCO, one of the largest collections of academic databases in the world, could not realistically be expected to have the papers

Box 1. Relationship codes

SIMPLE INTERACTIONS
> action directed at the communicatee(s)
< action directed at the communicator(s)
>< interaction
>//< parallel interaction
-> outcome ≈> negative outcome
÷ antagonistic relation
<÷> argument, disagreement
>+< agreement, reciprocity
} modification of component
C} cultural modification
M} & F} gender modification
@ modularity
- Verbal (words) w@txt
- Visual v@txt
- Musical m@txt
- gestural g@txt
- gustatory t@txt
- verbal & visual wv@text
- verbal, visual, &musical wvm@txt
- and other combinations

-- unidentified relation
+ elements combined in effecting an element ••• duration of process
[] have to be considered as an "interactive set" a micro system
[[ ]] a system
Bold element emphasized in model

Advanced Notations
>> power (commanding)
÷> deception
< responsive force
<÷ resistance
>+ expectation
^ expectation negated
~~ dysfunctional group
{} {} mutual modification—e.g., {C} tor} <>{Ctee}
} { tension between
?txt•••>txt! problem (?) solving (!) process
><txt•••>txt!>< collaborative problem solving process
txt> locutionary force
>tee perlocutionary force

already in its archives coded. However, if an organization such as EBSCO included in its protocols for submitting papers to the various journals it indexes the requirement that the research submitted for publication be coded in terms of this or a similar notational system, the most recently published papers would be able to be retrieved with more precision and flexibility than in the past. Moreover, as publications accumulated in the various EBSCO

Box 2. How proposed markup language identifies similar conceptions of communicating

[Tors>]TXT>>Ptx>tees	Maxwell McCombs & Donald Shaw	Agenda Setting Theory
The conception of agenda setting is a communication in which communicators "tors" functioning as a group "[tors]" construct a public text "TXT" so that it commands/directs ">>" the cognitive frameworks "ptx>" of communicatees "tees."		
[Tors>]TXT>>Ptx>tees	Todd Gitlin	Framing Theory
The coding of the conceptions of agenda setting and framing reflect their similarity.		
{C-ptx>}tor>txt<+>{tee<ptx-C}	Stella Ting-Toomey	Face Negotiation Theory
The conception of face negotiation involves mutual modification "{} {}" of cultural codes "C-ptx" directed ">" by the communicator "tor" in the construction of the text "txt" in a reciprocal "<+>" communication where the communicatee "tee" directs "<" cultural codes "ptx-C" toward the communicator.		

accessible databases, the usability of the archive would increase rapidly and make contemporary research, often lost to the vastness of research databases, more accessible and easier to find for researchers and students seeking research on specific contexts of use.

Given the circumstance that most papers submitted to journals or stored in databases are digital, it would not be difficult to develop software which volunteer or professional coders could use to markup previously published research. Starting with canonical or highly cited works, graduate students conducting research for dissertations or faculty preparing to teach theories in class could code those publications relevant to their own work. Considering the relatively minimal time commitment needed to code a single publication it should be easy for large numbers of previously published work to be coded by researchers leveraging them in their own projects. It would be possible for research database providers to develop simple systems for entering, tracking, and searching this coding system useful for marking up or locating research published in digital locations (as most are now).

We have been experimenting with programs that automate the entry of data from digital copies of publications into a lexicon as part of a Concept Toolkit at a website dedicated to Conceptual Logistics employed as a sociotechnical system for leveraging the computational power of web servers and the human power of the socially constructed Web. The experimental program set begins with a digital copy of an article, breaks it down into clauses, organizes the clauses in terms of categories based on the core functions we have described and produces rough drafts of the various entry fields, in particular the entries "conception," "embedded concepts" (in the main conception), "contexts of use," and "influences." The experimental program set also automatically produces the coding we have described. A contributor armed with our software, still under development, would only have to evaluate the rough draft of the text designated for specific entry fields and revise the draft accordingly with attention to its style:

- **Standards and Bodies:** The normal progression for a markup language, especially one that can be generalized to multiple cases, is to go from inception, through development and testing, and eventually on as a draft proposal to either a standards body or an organization from a field in which the markup could be used. The initial impulse is to take it to a large, internationally recognized body like the World Wide Web Consortium (W3C) or International Organization for Standardization (ISO). In reality this draft is neither ready for a

similar body or readily appropriate. The ISO likely would not find this draft useful or broad enough to merit its consideration.

Instead, much as the W3C is a focused body handling a specific domain of standards we envision either the creation of a Research Markup Standards Group (RMSG) that could administer a standardized markup system applicable across research domains or for major bodies within those fields (such as the International Communication Association or Institute of Electrical and Electronics Engineers) to take on the standardization of the markup language for use within their own fields. The former option appeals to notions of academic standards and wide-spread adoption but may not provide a sufficiently tailored markup language across fields. The latter approach would likely yield a number of highly focused (and potentially very useful) markup systems but could also spawn an unyielding array of competing and incompatible systems from research field to research field.

HOW THE PROPOSED INFORMATION SYSTEM WOULD WORK IN DATA RETRIEVAL

Since the coding describes conceptions with respect to the nine functions of communicating events, conceptions about any function or combination of functions can be retrieved. For example, a search for "fram*" and "S}txt" would retrieve all of the contexts of use for specific conceptions of "frame" and "framing." In cases where more than one article features a research conception, this approach reveals the changes made by comparing the conceptions to each other. The comparisons can be made in instances where the same researcher uses a particular term in several articles. For example, Robert Entman uses the conception of framing in numerous publications (e.g. 1991, 1993, 2003). A search for "Entman" + "txt" + "framing" (together with a specified word range) would retrieve all of the texts in which he uses the concept of framing which can then be compared. Given a large enough collection of coded texts even the comparison could potentially be done systematically by software:

- **Identifying Changes in the Focus of Communication Research:** Since the information system we are developing links the notational description of research projects to their dates of publication, it would be possible to track changes in the theoretical focus of communication research. For example, intra-personal communication (likely coded as "tor><tor") virtually disappeared after the 1980s. On the other hand, the advent of the Internet increased interest not only in the modalities of communication (for example, "wva@txt") but also in intercultural communication (for example, "C}tor><txt><tee{C");
- **Surveying the Proposed Range of Relations in Communicating Events:** If a student or researcher wanted to know what types of interpersonal relations have been researched, she could search for "tor*txt*tee" (the asterisk serving as a wild card) and a list of entries would show the research that has been done on this relationship. The search, of course, could be easily narrowed to a specific type of interpersonal relationship, for example, by searching for "C}tor><wv@txt><tee{C" to identify multimodal intercultural relationships which are characteristically Internet based;
- **Visualizing Related Concepts:** It is possible to format searches and results in concept webs. This would show how conceptions are networked semantically, a valuable resource for students. Similarly, it is possible to format searches in a timeline. For example, it would be possible to create a chronology for the history of communication research, showing the expansion and contraction of focus as well as the key figures in those changes. Both of these technologies exist as open source tools in

general, and we have begun to develop these approaches as part of the earlier referenced Concept Toolkit.

CONCLUSION

We began this proposal pointing to the difficulties students, researchers, and teachers have in finding the relationships among communication theories given the idiosyncratic naming practices used to identify research projects. Using the 9 Dot Puzzle as a metaphor for the importance of re-conceptualizing rules we proposed the creation of a notation system that powers a conceptual markup language for describing and locating the various knowledges locked up in the vast databases of academic research. To show how this notation system might work we defined a functional model of communication and developed a notation system for coding the components across contexts of use. Finally we defined and provided examples for our proposed notation system at work to illustrate its potential utility.

We, as communication researchers, focused this method of sociotechnical knowledge management on the field of human communication, typically referred to as Communication Studies in contemporary universities. However, we offer this theoretical and technical approach as a model that can be applied, in principle, to all fields of academic inquiry where discourse domains are used to organize published research. As research discourses grow, any field would benefit from the ability to systematically analyze and traverse the theories, embedded concepts, and contexts of use that structure the analytical landscape of academic research. The coding schema might vary across research domains, but the underlying principle of a markup language facilitating the construction of, and interaction with, published (and potentially even upcoming) work would be well worth the effort needed to construct them. This proposal shows how the present retrieval difficulties can be significantly ameliorated by applying the proposed notational system to modern communication research using existing technology.

REFERENCES

Austin, J. L. (1962). *How to do things with words*. Cambridge, MA: Harvard University Press.

Berlo, D. K. (1960). *The process of comnunication*. New York, NY: Holt, Rinehart and Winston.

Blumer, H. (1969). *Symbolic interactionism: Perspective and method*. Englewood Cliffs, NJ: Prentice-Hall.

Brown, G., & Yule, G. (1987). *Discourse analysis*. Cambridge University Press.

Burke, K. (1966). *Language and symbolic action: Essays on lift, literature, and methods*. Berkeley, CA: University of California.

Carlson, G., & Sosnoski, J. (2013). Infectious textbooks. In A. Whelan, C. Moore, & R. Walker (Eds.), *Zombies in the academy: Living death in higher education*. Intellect Press.

Charlton, W. (Ed.). (1970). *Aristotle's physics: Books 1 and 2*. Clarendon P.

Engler, B. (1991). Textualization. In R. D. Sell (Ed.), *Literary pragmatics* (pp. 179–189). Routledge.

Entman, R. (1991). Framing U.S. coverage of international news: Contrasts in narratives of the Korean and Iranian airline incidents. *The Journal of Communication*, *51*(4).

Entman, R. (1993). Framing: Toward clarification of a fractured paradigm. *The Journal of Communication*, *43*, 51–58. doi:10.1111/j.1460-2466.1993.tb01304.x

Entman, R. (2003). Cascading activation: Contesting the White House's frame after 9/11. *Political Communication*, *20*, 415–432. doi:10.1080/10584600390244176

Fillmore, C. J. (2006). Frame semantics. In D. Geeraerts (Ed.), *Cognitive linguistics: Basic readings* (pp. 373–400). Berlin, Germany: Mouton de Gruyter. doi:10.1515/9783110199901.373

Gee, J. P. (1999). *An introduction to discourse analysis: Theory and method*. Routledge.

Georgakopoulou, A. (2004). *Discourse analysis: An introduction* (2nd ed.). Edinburgh, UK: Edinburgh University Press. doi:10.3366/edinburgh/9780748620456.001.0001

Griffin, E. (2012). *A first look at communication theory* (6th ed.). Boston, MA: McGraw Hill.

Halliday, M. A. K. (2009). Context of culture and of situation. Webster, J. J. (Ed.), The essential Halliday. London, UK: Continuum.

Halliday, M. A. K., & Kress, G. R. (1976). *Halliday: System and function in language: Selected papers*. Oxford University Press.

Infante, D. A., Rancer, A. S., & Womack, D. F. (2003). *Building communication theory*. Long Grove, IL: Waveland Press.

Kuhn, T. (1965). *The structure of scientific revolutions*. Chicago, IL: U of Chicago P.

Kuhn, T. S., Conant, J., & Haugeland, J. (2000). *The road since structure: Philosophical essays, 1970-1993, with an autobiographical interview*. Chicago, IL: University of Chicago Press.

Langacker, R. W. (2002). *Foundations of cognitive grammar*. Stanford, CA: Stanford University Press.

Littlejohn, S. W., & Foss, K. A. (2011). *Theories of human communication* (10th ed.). Long Grove, IL: Waveland Press.

Lyons, J. (1977). *Semantics*. Cambridge University Press.

Miller, K. (2005). *Communication theories: Perspectives, provesses, and contexts*. Boston, MA: McGraw Hill.

Mumford, E. (1989). *XSEL's progress: The continuing journey of an expert system*. Wiley.

Mumford, E. (1996). *Systems design: Ethical tools for ethical change*. Hampshire, UK: MacMillan.

Mumford, E. (2003). *Redesigning human systems*. Hershey, PA: Idea Publishing Group.

Philipsen, G. (1997). A theory of speech codes. In P. Gerry, & L. A. Terrance (Eds.), *Developing communication theories* (pp. 119–156). Albany, NY: SUNY Press.

Reisberg, D. (2010). *Cognition: Exploring the science of the mind* (4th ed.). New York, NY: W. W. Norton.

Schramm, W. (1971). How communication works. In J. A. DeVito (Ed.), *Communicaton: Concepts and processes* (pp. 12–21). Englewood Cliffs, NJ: Prentice-Hall.

Schramm, W., Chaffee, S. H., & Rogers, E. M. (1997). The beginnings of communication study in America: A personal memoir. *Sage (Atlanta, Ga.)*.

Searle, J. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press. doi:10.1017/CBO9781139173438

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, UK: Cambridge UP. doi:10.1017/CBO9781139173438

Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

Shapere, D. (1977). Scientific theories and their domains. In F. Suppe (Ed.), *The structure of scientific theories* (pp. 518–565). Urbana, IL: U of Illinois P.

Tulving, E., & Craik, F. I. M. (Eds.). (2000). *The Oxford handbook of memory*. Oxford, UK: Oxford UP.

Vosniadou, S. (Ed.). (2008). *International handbook of research on conceptual change*. Routledge.

Werth, P. (1999). *Text worlds: Representing conceptual space in discourse*. London, UK: Longman.

West, R., & Turner, L. H. (2010). *Introducing communication theory: Analysis and application*. Boston, MA: McGraw Hill.

Wittgenstein, L. (1953). Philosophical investigations. In G. E. M. Anscombe (Trans.), Basil Blackwell.