Title
Exploded Conncetions: Unchunking Schematic Knowledge

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 4(0)

Author
Small, Stephen L.

Publication Date
1982

Peer reviewed

Exploded Connections:

Unchunking Schematic Knowledge

Steven L. Small
Department of Computer Science
The University of Rochester
Rochester, New York 14627

Background

It has been understood for some time that the organization of knowledge into event schemata and visual schemata can aid significantly in the inference-making process. If we know that we are in a typical room, to use Minsky's example [1974], then we expect to see windows, walls that are perpendicular to a ceiling, etc. If we know that we are at a restaurant, to use Schank's example [1975], then we can expect to be seated by a maitre d'hotel, to be given menus, etc. By classifying situations according to a small collection of schematic situations, a wide variety of inferences become immmediately clear and simple.

This same kind of schematic reasoning constitutes the heart of several well-known theories of low-level comprehension, especially by Schank [1972], Wilks [1973], etc. By classifying linguistic clauses into a small number of semantic categories, such as physical transfers of location (PTRANS), propelling of objects (PROPEL), etc., a number of inferences are straightforward. Certain kinds of paraphrase are simple: "buying" and "selling" are represented in almost the same way; "running," "walking," and "biking" have much in common in their semantic representations. The schema for abstract transfers of possession (ATRANS) leads us to expect exchange of one thing for another from one person to another. If any of these are not specifically specified, they can be inferred easily. Further, such an abstract transfer probably took place because one person wanted to own something that had been owned by someone else, the other person probably didn't want it so much anymore, and similar kinds of simple inferences. The MARGIE system [Schank et al, 1973] exhibited very impressive behavior without using much more than schematic inferences based on the semantic representation scheme of Conceptual Dependency (CD).

Unchunking Schemata

In this short paper, we suggest a framework for the study of schematic aspects of natural language comprehension. Specifically, we pursue the tack of Feldman [1976] in preferring the dynamic rather than static chunking of knowledge; the use of parallel [illegible] diffuse knowledge representation facilitates that goal. The approach draws from previous work in schematic representation and reasoning [Minsky, 1974; Schank, 1975], spreading activation [Quillian, 1968], parsing [Small, 1980; Marcus, 1979], speech recognition [Lowerre, 1976], psycholinguistics [Dell, 1980; McClelland and Rumelhart, 1980], and computer vision [Ballard, 1981; Marr, 1978]. By decomposing schematic knowledge into diffuse units and by studying the way these facets of knowledge are connected (inferentially), we expect to show important results in several areas:

- how a language comprehension system can maintain diffuse loci of control (hypotheses) simultaneously, but still come to a decision when required;
- how to obtain schematic reasoning (and expectations) from distributed units not a priori committed to representing unique static situations;
- how to merge schematic inference mechanisms from the top-down (e.g., scripts) and from the bottom-up (e.g., case frames); and
- how to relate experimental psychological data (e.g., reaction times on normals and aphasics) to computer models.

The modelling effort employs an architecture significantly different from the typical computer and closer to that of the human brain. We use a particular spreading activation or active semantic network scheme, called *connectionism*, which consists of a massive number of appropriately connected computing units that communicate through weighted levels of excitation and inhibition [Feldman and Ballard, 1982]. While such an architecture does not solve any problems per se, we believe that a number of questions become easier to set forth and more straightforward to solve. This paper intends only to suggest the directions of our current research in addressing several language comprehension issues from the new perspective.

Some Main Issues

In particular, we show how a number of classical problems in the theory of schemata might be approached in a new way. Three principal issues are discussed: (1) Comprehension takes place on a number of interacting levels of processing; (2) multiple hypotheses are simultaneously maintained at a number of diffuse processing loci; and (3) context affects processing in both top-down and bottom-up directions. Experimental psychologists are beginning to understand these issues through reaction time data: McClelland and Rumelhart [1980] have identified two processing levels; Dell [1980] presents data suggesting interactions within the phonemic level; Swinney [1979] shows ways in which context *does not* affect processing; Seidenberg et al [1980] illustrate an entire time course for processing at the lexical level; and Samuel et al [1982] present data suggesting the mechanisms of letter processing and the word superiority effect.

Multiple Levels of Comprehension

While it is sometimes the case that a language understander needs to know the primitive schematic actions that compose a more complex action, often he does not. The relevant information carried by particular words and expressions is precisely that information that aids the hearer to understand the intended meaning of the speaker. This always takes place in some context and cannot be separated from it. In a dialogue, a hearer must interpret the words and expressions in light of the communicative goals of the speaker; in a story, a reader must serve to connect new fragments of text with the existing interpretation of story structure. Further, general knowledge about the world must be applied where needed to the comprehension process (even at the level of individual words and phrases), and the story or goal structures constructed must be constrained by it.

It seems unusual to consider certain actions in terms of their decompositions (in the sense of CD) into structures of primitive units. There are very few contexts in which the sentence "Rick kissed Joanie" would be best understood by focusing on the (nonetheless valid) fact that "Rick moved his head to in front of the head of Joanie so that they were both facing each other, and then puckered his lips and touched them to Joanie." This long description must be represented as the algorithmic (functional) concept underlying kissing. This description would be required to understand the sentence "Joanie caught Rick's cold" occurring next. It would certainly not help in understanding the sentence "Joanie bought herself a new blouse." The understanding of this second conceivable utterance requires an entirely different set of relationships concerning kissing.

Multiple Simultaneous Loci of Control

Thus, we need at least two different kinds of associations of kissing to understand sentences in which it is a central action. When hearing such a sentence, both of these kinds of associations are activated, and either one can be relevant to understanding what comes next. Furthermore, the context previous to the kissing action could serve to make one or the other of these associations the prominent one [illegible] the understanding. The sentence "Rick didn't care about the flu" would facilitate understanding the next sentence in a way more heavily weighed toward the algorithmic association of kissing than the emotional one. Likewise, the preceding sentence "Rick felt strongly affected" should facilitate the other associations. An active processing network works through simultanous activity in many processing locations, permitting a cognitive model to avoid irrevocable all-or-none decisions in favor of a more continuous approach. This leads to plausible explanations of subsequent context effects, including puns.

Context Effects on Comprehension

In building a computer model of language comprehension, we must consider these phenomena. The context preceding an utterance must serve to favor certain interpretations over others. As in the example presented here, the competing interpretations need not be incompatible; the context must simply facilitate the comprehension of subsequent utterances by focusing on one level of interpretation over others. It should take longer for a hearer to understand how Rick could get the flu from Joanie if conditioned to focus on their emotional involvement, than to understand the same utterance after contextual conditioning to focus on the mechanics of kissing. The results of Seidenberg et al [1980] suggest that analogous contextual effects hold with respect to lexical access.

The computer model should make accessible all levels of interpretation of utterances, but should not make everything as easily accessible as everything else. When knowledge of the mechanics of kissing are required to understand a fragment of text, it must be available. If the text is about some romantic relationship, this knowledge would usually be an obstacle, rather than a help, to understanding. In such a case, it should be available if needed (though perhaps slowly), but mostly it should not be involved.

The Exploded Connection Scheme

We propose a uniform representation scheme for both high- and low-level language processing that shares some of the flavor of the schematic approaches, but which incorporates flexibility through three methods:

(1) The use of incredibly large numbers (and wide variety) of schematic situations (*units*);

(2) A focus on the relationships among these situations rather than on the situations themselves (*connections*), and

(3) The use of numerical potentials to weigh (comparatively) the relevance of any particular schematic situation to the data (*activation levels*).

We call the approach an *Exploded Connection Scheme* (ECS), and are using it to build a unified theory of low- and high-level language comprehension. The elemental units of ECS encompass the gamut of traditional elements of comprehension models, from phonemes and morphemes to cases and semantic primitives to concepts and event sequences. We believe that there are large numbers of each kind of unit, and that reasoning takes place through the richness of the unit vocabulary and the connections among the individuals. Some of our current ideas on the organization of these exploded units can be found in [Cottrell, 1982].

What we are arguing for is a highly diffuse active representation of knowledge and its processing. Traditional models [Schank et al, 1973; Small, 1980] represent the meaning of utterances in single, large, complex structures of some small number of primitive elements. Large processes then manipulate this knowledge, encoding and decoding the large symbol structures. Alternatively, we are suggesting representing meaning in a very large number of (exploded) active processing units, which compute activation as a function of incoming weighted excitation and inhibition. The scheme focusses on the complex interactions among diffuse knowledge units and reduces the complexity of individual processes. Such processing units that do not manipulate complex symbol structures can interact frequently and tightly.

The Pair Principle

Each unit of a particular type triggers activation in other similar units that are likely to come next in meaningful speech. This happens at every level of processing, within the level itself, and is called the *pair principle*. The principle states that every element of knowledge triggers other elements (of the same kind) that are likely to succeed the given one (temporally or inferentially) in meaningful speech. Dell [1980] shows a model for speech production in which connections according to this pair principle lead to plausible explanations of experimental results in production of speech errors. By adding connections between levels, activity spreads through the network in a way that leads to predictions at every level of processing about what is coming next. The spreading activation model of McClelland and Rumelhart [1980] shows such predictions through grapheme/lexeme interactions. The results of Samuel [1980] on word superiority also support this view.

Examples of the pair principle can be seen at every conceivable level of language processing. A phoneme unit activates those other phoneme units that can follow the given one under the rules of the phonological system of a particular language (or the phonologic rules known to the hearer). A high-level activity unit activates the units for other high-level activities that can reasonably come next under the rules of cultural behavior of a society (or analogously, those rules known to the understander). An action unit increases the potentials of type units that represent the kinds of things that could be the case fillers of that particular action. The pair principle underlies the way we go about connecting units together within particular levels of the comprehension model.

Schemata

At the level of high-level activities the pairwise connections of units might seem less obvious than at the level of phonemes. The sound pattern of languages are well-known, but not so the cultural regularities. Further, the differences among individuals may seem greater when it comes to their expectations about events in the world as opposed to their use of sounds. We contend that this is not so. The restaurant script of Schank [1975] constitutes a good example of the expectations of people from our culture about the high-level activities involved in eating at a restaurant. Two fundamental problems are known to exist with scripts: (1) how do you know when one is relevant; and (2) how do you use information from one script in understanding activities in another?

Schank [1979] has begun to address these questions in his recent work on MOPs, or memory organization packets. In our way of viewing language comprehension, the Yale group has shifted slightly its emphasis: they are increasing the number and nature of the schematic situations they recognize and they are focusing a bit more on the connections among situations. Their restaurant script is now connected to other schemata representing the general notions of visiting a business establishment, preparing and eating a meal, out-of-house social activities, etc. We agree with this shift, and push it to its logical conclusion, as enumerated above in our three representation methods.

A conceivable pair matrix for several example high-level activities of restaurant-going is shown in Figure 1. We can envision additional pair matrices for each activity found on the right-hand side of the one shown--the entire set of connections being quite large. During the comprehension process, the activation of one activity unit causes concomitant model activity in those that follow it. Further, this pairwise triggering does not stop at the units that are only one connection away, but continues (at a smaller-valued potential) for a good distance, activating a large number of units until the ever-decreasing value is no longer significant.

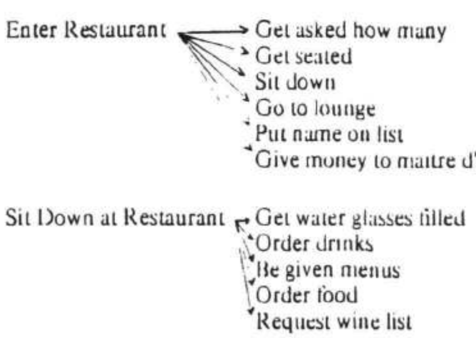

Figure 1. *Activity Pair Relations*

Hierarchy

One kind of inference has always been a problem for schema-based models of comprehension based on static chunks of knowledge and a single processing focus. In the birthday party scenario of Charniak [1972], what does the model do if something happens not directly related to parties? Likewise, in his later exploration of the world of painting [1977], how does a schema driven model understand an event in a story that has nothing to do with painting? An attempted solution within the single process approach has been to use some sort of stack mechanism for the schemata, such that one context gets pushed (to use the computer metaphor) and another takes over. Again, the problem arises: which schema to activate when the previous one is pushed? And when is it popped?

The solution to this classic problem within the Exploded Connection Scheme involves connections among hierarchical levels of active knowledge units in the model. Research in semantic networks has led to interesting epistemologies and computer representations. The work of Fahlman [1979] in particular shows how the different levels of description might be related to each other in our own scheme. The main difference between his NETL approach and ECS centers on the elimination of a central controller in favor of multiple competing processing loci, each with a dynamic activation (confidence) level.

The pair matrix that shows some of the activities involved in going to a restaurant and the activities likely to follow them in everyday circumstances in our culture (Fig. 1) seems specific to that overall activity, i.e., restaurant going. The events listed are exploded, in the sense that they describe "entering a restaurant" and "being seated at a restaurant," rather than "walking" and "sitting down." This explosion means that there are a large number of different units all representing the same kind of activity in different contexts. If we leave things as such, there are many problems of schema-based models that will cause trouble in our scheme. How can we reason that "sitting down in a restaurant" could lead next to a "knee spasm," for example? Whether we represent the knowledge as "the sitting down" action of the "restaurant schema" or as the "sitting down in a restaurant" unit in a connected network, the inference is not possible (unless "knee spasms" are explicitly linked to the "sitting down in a restaurant" unit, an unacceptable solution).

Our solution to this problem is to connect every unit in the exploded network with a number of units that represent the same event less specifically. The method we use, called the ***hierarchy principle***, involves representing events in an ever more specific hierarchy, from completely context-free actions, such as ingesting, to very particular ones, such as "eating squid at a Spanish restaurant in Georgetown." A small set containing a few intermediate kinds of eating is illustrated in Figure 2. While it might seem like there are far too many units in this hierarchy to be plausible as a representation of knowledge for comprehension, it is important to realize that: (1) most of the possible units do not exist in each individual; (2) the hierarchy is a tangled one; and (3) only the units at the highest levels are fairly fixed in the nature of what they represent.

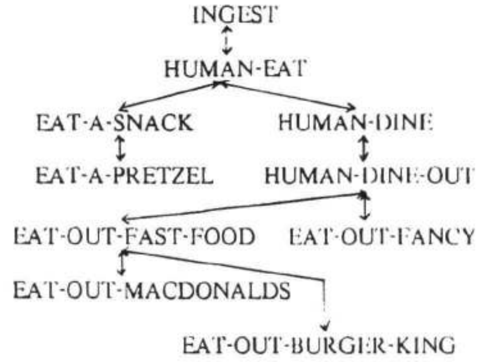

Figure 2. *Hierarchical Activities*

Connections and Spreading Activation

Let us refer to the pairwise connections that are based on the temporal order of processing (e.g., phonemes, events) as *temporal connections*, those based on mundane inference as (*mundane*) *inference connections*, and those in the hierarchy as *hierarchical connections*. Sometimes it is convenient to call connections of the first two kinds *follow connections*. The combination of spreading activation along these different pathways provides an answer to the problem with schema-based comprehension mentioned previously. The temporal and inferential order of events now triggers activation in event units in two dimensions, leading in the horizontal direction to expectation of specific schematic events, and in the vertical direction to non-schematic events of a less-specific nature. Note that the vertical activation causes activation horizontally among these more general activity units.

The restaurant-going example can be used to illustrate the nature of this activation. When the "being seated at a restaurant" unit becomes active, a number of event units along the follow connection pathways are also activated. These include such things as "being given a menu," "asking for a wine list," etc., as shown in the pair matrix of Figure 1. That activation in turn causes additional activation at a lower level along the next set of follow connections, and so on, until the ever decreasing activation has become essentially zero.

Simultaneously, however, activation proceeds along the hierarchical connections as well, likewise decreasing for each new radius of connection. Of course, each event unit in the hierarchy has both hierarchical and temporal connections, and activation from hierarchical paths proceeds out along all connections, regardless of type, thus creating a new set of event pair expectations. The way that "being seated in a restaurant" can naturally lead to the comprehension of a "knee spasm" through a combination of hierarchical and temporal pathways is shown in Figure 3. By the decreasing activation idea, the potential (activation level) of "knee spasm" should be significantly less than that for things like "looking at the menu," but that is perfectly consistent with our thoughts on how context strongly affects perception of new inputs.

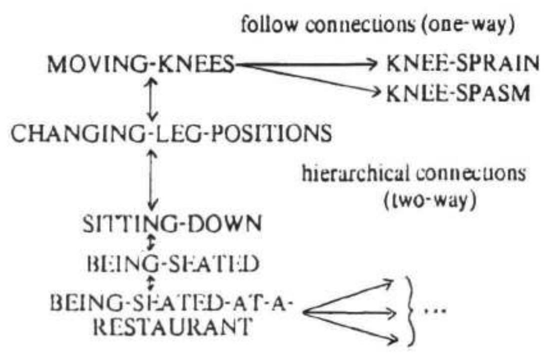

Figure 3. *Schema Interaction*

Furthermore, in cases when stimuli could be interpreted in more than one way, this scheme leads to hypotheses (experimentally testable) about the preferred interpretation. The role of perception is also important in the scheme, since the activation level of units depends on inputs along all dimensions of connectivity in the network. The potential of a unit can be changed by direct perceptual stimulation, stimulation from above or below in the isa hierarchy, from previous events along the follow pathways, or from other sources yet to be identified. The potential represents in a uniform way the stimulation provided by a combination of all incoming connections. The construction of our model involves identifying the nature of connections and units (which we are now doing) and the nature of the combination rules for each kind of unit.

Summary and Conclusions

The research program we are commencing -- the construction of a computer model of human language comprehension -- represents an interdisciplinary effort in cognitive science. The plan involves connecting up a large number of neuron-level computing units to process cohesive text. Empirical constraints on the organization of the active network consist of processing evidence from psychology, physiological evidence about the brain, and computational plausibility. Thus far, we have made some preliminary studies in parsing and schematic reasoning, and have a working network simulator that has been applied successfully to some simple problems in high level constraint relaxation.

In this paper, several issues regarding the organization of schematic knowledge for language comprehension have been described within the connectionist framework. We have suggested mechanisms for (1) obtaining schematic reasoning from diffuse computing units; (2) merging top down and bottom-up control in schematic reasoning; (3) maintaining diffuse loci of control yet coordinating global behaviors; and (4) directly relating psychological evidence to computational models in cognitive science. The results presented are certainly in a preliminary state, but are leading to interesting simulations and valuable collaborative work.

References

Ballard, D.H., "Parameter networks: Towards a theory of low-level vision," *Proc*, 7th IJCAI, Vancouver, B.C., August 1981.

Charniak, E., "Ms. Malaprop, a language comprehension program," *Proc.*, 5th IJCAI, 1977.

Charniak, E., "Toward a model of children's story comprehension," AI Memo 266 AI Lab, MIT 1972.

Cottrel, G., "Toward Connectionist Parsing", Technical Report, Department of Computer Science, University of Rochester, 1982 (to appear).

Dell, G.S., "Phonological and Lexical Encoding in Speech Production", Ph.D. dissertation, Department of Psychology, University of Toronto, 1980.

Fahlman, S.E. *NETL, A System for Representing and Using Real Knowledge.* Boston, MA: MIT Press, 1979.

Fahlman, S.E., D.S. Touretzky, and W. van Roggen, "Cancellation in a parallel semantic network," TR, Computer Science Dept., Carnegie-Mellon U., 1981.

Feldman, J.A., "Bad-mouthing frames," *Proc.*, TINLAP, 1976.

Feldman, J.A. and D.H. Ballard, "Connectionist models and their properties," to appear in *Cognitive Science 6*, 1982.

Hinton, G.E., Review of S.E. Fahlman, *NETL, A System for Representing and Using Real-World Knowledge. AISB Quarterly,* 42-43, Winter/Spring 1981/82.

Lowerre, B.T., "The Harpy Speech Recognition System", Ph.D. dissertation, Department of Computer Science, Carnegie-Mellon University, 1976.

Marcus, M.P., "An Overview of a Theory of Syntactic

Recognition for Natural Language." AI Memo 531, MIT Artificial Inteligence Laboratory, 1979.

Marr, D., "Representing visual information," in A.R. Hanson and E.M. Riseman (Eds). *Computer Vision Systems.* NY: Academic Press, 1978.

McClelland, J.L. and D.E. Rumelhart. "An interactive activation model of the effect of context in perception: Part 1," Report 8002, Center for Human Information Processing, U. California, San Diego, May 1980.

Minsky, M., "A framework for representing knowledge," in Winston (Ed). *The Psychology of Computer Vision.* McGraw-Hill, 1974.

Minsky, M., "K-Lines: A theory of memory," *Cognitive Science 4*, 2, 117-133, 1980.

Norman, D.A., "A psychologist views human processing: Human errors and other phenomena suggest processing mechanisms," *Proc.*, 7th IJCAI, 1097-1101, Vancouver, B.C., August 1981.

Quillian, M.R., "Semantic memory," in Minsky (Ed). *Semantic Information Processing.* Boston, MA: MIT Press, 1968.

Rieger, C.J., "The Importance of Multiple Choice", *Proceedings TINLAP-2*, Urbana, Illinois, 1978.

Rosenfeld, A., R.A. Hummel and S.W. Zucker "Scene labelling by relaxation operations," *IEEE Trans. SMC 6*, 1976.

Samuel, A.G., J.P.H. van Santen, and J.C. Johnston, "Length effects in word perception: We is better than I but worse than you or them," *J Experimental Psychology: Human Perception and Performance 8*, 1, 91-105, 1982.

Schank, R.C., "Conceptual dependency: A theory of natural language understanding," *Cognitive Psychology 3*, 4, 1972.

Schank, R.C., N. Goldman, C. Rieger, and C. Riesbeck, "MARGIE: Memory, analysis, response, generation, and inference on English," *Proc.*, 3rd IJCAI, 1973.

Schank, R.C., and R.P. Abelson, "Scripts, plans, and knowledge," *Proc.*, 4th IJCAI, 1975.

Schank, Roger C., "Reminding and memory organization: An introduction to MOPs," Research Report 170, Dept. of Computer Science, Yale U., 1979.

Seidenberg, M.S., M.K. Tanenhaus, and J.M. Leiman, "The Time Course of Lexical Ambiguity Resolution in Context", Technical Report #164, Center for the Study of Reading, University of Illinois, 1980.

Small, S.L., "Word Expert Parsing: A Theory of Distributed Word-Based Natural Language Understanding," Ph.D. dissertation and TR 954, Department of Computer Science, U. Maryland, 1980.

Swinney, D.A., "Lexical access during sentence comprehension: (Re)consideration of context effects," *J Verbal Learning and Verbal Behavior 18*, 645-659, 1979.

Tanenhaus, M.K. and J.M. Leiman, "Evidence for multiple stages in the processing of ambiguous words in syntactic contexts," *J Verbal Learning and Verbal Behavior 18*, 427-440, 1979.

Wilks, Y., "Preference semantics," AI Memo 206, AI Lab, Stanford U, 1973.

Wilks, Y., "Some thoughts on procedural semantics," Cognitive Studies Centre Report 1, U. Essex, 1980.