Project Description

The goal of this project is to develop a detailed model of the structures found at the semantic level of natural language. The project represents the final phase of research on the structural levels of language: phonetic, phonemic (Leinieks, 1965, 1975b), morphemic (Leinieks, 1964a, 1964b, 1967, 1975a), syntactic (Leinieks 1964a, 1964b, 1975a), and semantic. A flowchart of the language process developed on the basis of this research is given in Figure 1. I have also dealt with semantics from the point of view of literary structure (Leinieks, 1982, 1996). Furthermore, I have also taught these subjects in linguistics, language, and literature classes for some forty years now.

Language at its various levels is an information processing device. Like all information processing devices it operates on the principles of contrast and cycle. Contrast means the mutually exclusive occurrence of elements from a specific set of elements in a specified location. In a computer, for example, a 1 or a 0 in any given location. Cycle means the repeated use of a sequence of such locations in a specified order. The cycle makes the encoding of an unlimited number of messages possible. The principle of contrast and the principle of cycle also make it possible to determine what elements are used in a given information processing device and what are the relationships among them. In natural language five separate levels of structure can be distinguished on the basis of five observable cycles of increasing length and complexity: a phonetic level with the phone cycle, a phonemic level with the syllable cycle, a morphemic level with the word cycle, a semenic level with the sentence cycle, and a semantic level with the story cycle. Simultaneous or long elements become more prominent as we go from the morphemic, to the sememic, to the semantic level. Sememic (syntactic) structures consist to a large extent of simultaneous clause and phrase sememes. At the semantic level simultaneous elements are even more prominent. They represent the real-time presence of various elements to the comprehension process. Topics, for example, a simultaneous long elements. So are speech-act identifiers. Appropriate use of simultaneous elements greatly simplifies semantic representations and helps explain the process of language comprehension. Natural language, as it is configured, is a rather strange device. Semantic units have come to be connected with phonetic units through three intermediate stages. The whole process operates something like an automobile transmission. Very long cycles with very many elements are stepwise changed to very short cycles with few elements. These can then be transmitted in the form of sound.

The model of semantic structure proposed here is a discourse model. Its basic features are the following. Each paragraph contains a small number of simultaneous long-element arguments or topics. These are time and place specifiers and usually one or two persons or things. Other arguments (persons or things) occur in paragraphs for briefer periods of time. The arguments are linked to one another in a succession of time frames either directly or through semantic predicates. Each time frame contains elements related to the same reference time. The arguments and predicates are linked through a variety of semantic features, which also serve to select possible linkages. These features are not the familiar agent, patient, instrument, etc., which are in reality only disguised syntactic elements out of place at the semantic level. The various arguments, above all the topics, are linked to memory structures that can be searched rapidly. Common memory structures include whole-

part and possessor-possession relationships. They also include generalizes abstract representations of various items and processes.

The advantages of the model proposed here can be best illustrated by comparing it with the widely known and widely used model proposed by Kintsch (1974). Kintsch recognized that the semantic structures of language are the same as memory structures. The Kintsch model was further refined by van Dijk and Kintsch (1983). Three distinct structures were identified: surface form (= syntactic structures), text base (= semantic structures), and situation model (= visual memory structures, for example). The theory is lucidly discussed by Fletcher (1994).

The Kintsch model of semantic structure uses a propositional representation, as do many other semantic models (Allen, 1995, for example). The text base (= semantic structures) is made up of propositions, each consisting of a predicate and one of more arguments. Thus José *ate a sandwich* can be semantically represented by EAT (AGENT: JOSÉ, OBJECT: A SANDWICH). Slightly different conventions are used by different investigators. All such propositional semantic representations, however, have serious limitations. These arise from the fact that the propositional representations are not significantly different from syntactic representations. Thus simplified syntactic representation of the above sentence can be written as:

sentence		
statement		
subject	predicate	object
José	ate	a sandwich

The lines below the labels indicate their extent as long or simultaneous constituents. The information given by the propositional semantic representation is identical with that given by the syntactic representation. Propositional semantic representations are simply a rewriting of syntactic structures in a different notation. As such they fail to capture the specifically semantic features which they should represent. There is in general little agreement among various investigators about the exact structures of the text base and the situation model and about the relationship between the two.

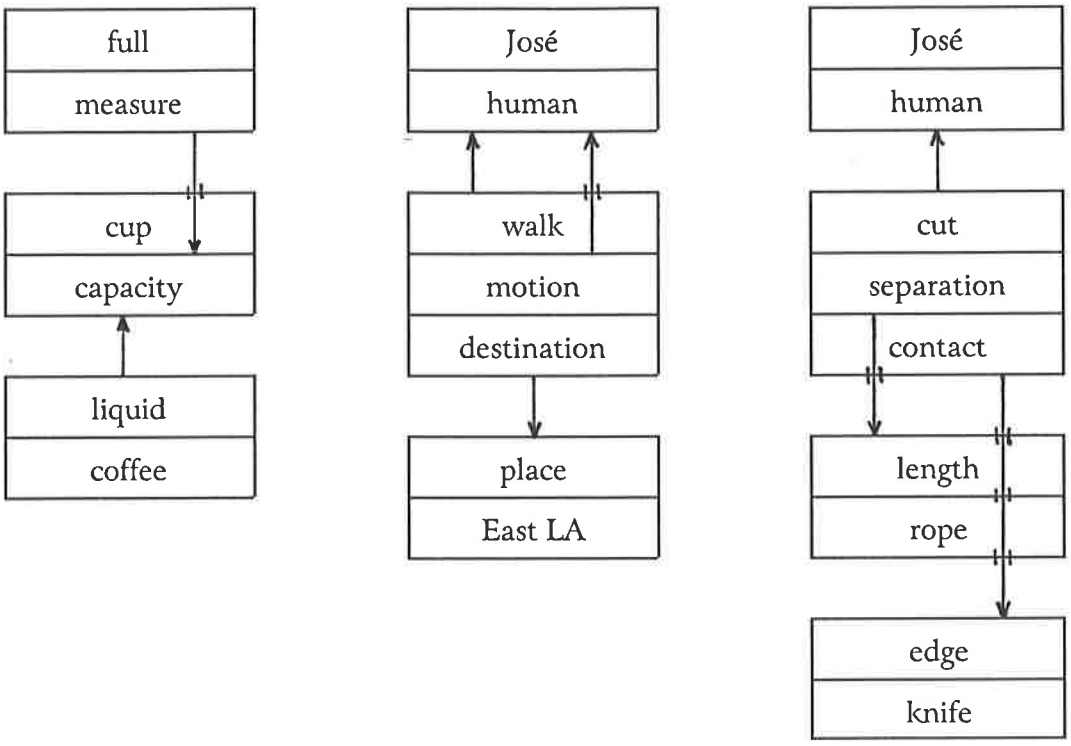
To produce a more accurate and complete semantic representation than the proposition model of Kintsch and others, at least the following things must be done.

1. A time axis, which indicates for how long various semantic elements are present to the comprehension process, must be introduced into the semantic representation. Let us consider the following example. *José walked over to the window and looked outside. A fight had erupted in the street. Bullets were flying. He turned away in disgust and went to the back of the house.* In this sample, consisting of three time frames, the topic José is present continually. In the second time frame he is part of the speech-act specifier. There are two place specifier elements, José's house and the street outside. The first extends over the first and third time frames; the second over the second time frame. Everything that is known about José, his house, and the street outside is connected to the respective elements and is readily accessible.

The time axis is also used to represent time anterior, simultaneous, and subsequent to the reference time in a time frame. Thus, in José, *who had arrived late, wanted to stay longer* the topic José is present continually throughout the time frame, while *arrived late* is present in a prior time, *wanted* at the reference time, and *stay longer* in subsequent time.

2. The conditions subsequent to an event predicate (= change predicate) must be represented explicitly in subsequent time. Thus José *broke the cup* automatically indicates *The cup was broken* in subsequent time and is present there to the comprehension process. This can again be done with long elements.

3. Semantic features that control the possible linkages between arguments and predicates or arguments and arguments must be specified in the semantic representation. These features are for the most part concrete and commonsensical. Thus, for example, *A cup of full of coffee, José walked to East LA,* and *José cut the rope with a knife* are linked as follows:



The minimal necessary number of these semantic features needs to be established.

4. Semantic representations are in many instances markedly different from syntactic representations. Thus semantic predicates may be encoded either by a syntactic predicate or a syntactic phrase. Thus *José arrived* and *José's arrival* have essentially the same semantic structure. Syntactic speech-act predicates along with their performers have a semantic modal function. They indicate long components simultaneous with the whole content of the speech act. Most importantly, topics, which are syntactically indicated by various anaphoric devices, are semantically single continuous long elements.

5. The simultaneous topic elements, the automatically generated subsequent conditions, and the extensive memory structures attached to the arguments provide at each point large amounts of material both immediately present and readily accessible by search to the comprehension process. Bridging inferences are facilitated by the fact that what appear as morphologically distinct but semantically related vocabulary items contain the same semantic feature. Thus *rain, downpour, drizzle, thunderstorm* will all contain some such features as

precipitation. Morphologically different parts of speech likewise contain the same semantic feature, as *heat* (noun), *hot* (adjective), *heat* (= make hot), for example. This greatly simplifies memory search processes.

6. Individual memory structures appear to contain both sense-data elements and semantic features. Thus, complicated spatial relationships, even abstract, generalized representations, are clearly stored in visual terms. They cannot be accurately and economically communicated verbally. It is possible, however, to assign as needed rather inexact semantic features to them, such as *above, behind, to the left*, etc. Other features in memory structures, such as *human, animal, thing*, or *male* and *female* are more economically encoded as semantic features.

These concepts of semantic structure need to be worked out in detail by means of an exhaustive analysis of a specific text. One of the problems with semantic theorizing is that it is frequently based on the analysis of carefully picked examples illustrative of an unusual problem. The analysis of an actual text provides a more complete picture and is a much better test of how well a semantic theory works. The texts proposed for analysis here are the writings of Julius Caesar, *Dē bellō Gallicō* and *Dē bellō cīvīlī*. The reasons for selecting these texts are the following. (1) I have already worked out the syntactic structure of Caesar and Cicero (Leinieks 1975a). (2) Latin is a language that indicates time relationships among various semantic elements very explicitly. This greatly facilitates the identification of time frames. (3) Latin morphology is highly systematic and transparent, which greatly facilitates the identification of semantic features. English morphology, by contrast, because of the varied sources of the vocabulary, is far more complex. (4) Latin has been studied from a linguistic point of view for a long time and much information about the language is already available. (5) It is easier to separate stylistic features from the minimal message carrying elements in a language which is not one's usual medium of communication.

References

James Allen (1995). *Natural language understanding*. Redwood City, CA: Benjamin/Cummings Publishing.

Charles R. Fletcher (1994). Levels of representation in memory for discourse. In Morton Ann Gernsbacher, editor. *Handbook of psycholinguistics*, 589-607. San Diego: Academic Press.

Walter Kintsch (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.

Valdis Leinieks (1964a). *Morphosyntax of the Homeric Greek Verb*. The Hague: Mouton.

Valdis Leinieks (1964b). The English predicate constituent. *Linguistics 3*, 30-49.

Valdis Leinieks (1965). The phonemic code of English. *Linguistics 14*, 23-32.

Valdis Leinieks (1967). The system of English suffixes. *Linguistics 29*, 80-104.

Valdis Leinieks (1975a). *The structure of Latin: An introductory text based on Caesar and Cicero*. New York: MSS Educational Publishing.

Valdis Leinieks (1975b). Defining the phoneme. *From meaning to sound: Papers from the 1974 Mid-American Linguistics Conference*, 249-253. Lincoln, NE.

Valdis Leinieks (1982). *The plays of Sophokles*. Amsterdam: Grüner.

Valdis Leinieks (1996). *The city of Dionysos*. Stuttgart: Teubner.

Teun Adrianus van Dijk and Walter Kintsch (1983). *Strategies of discourse comprehension*. New York: Academic Press.

The Language Process

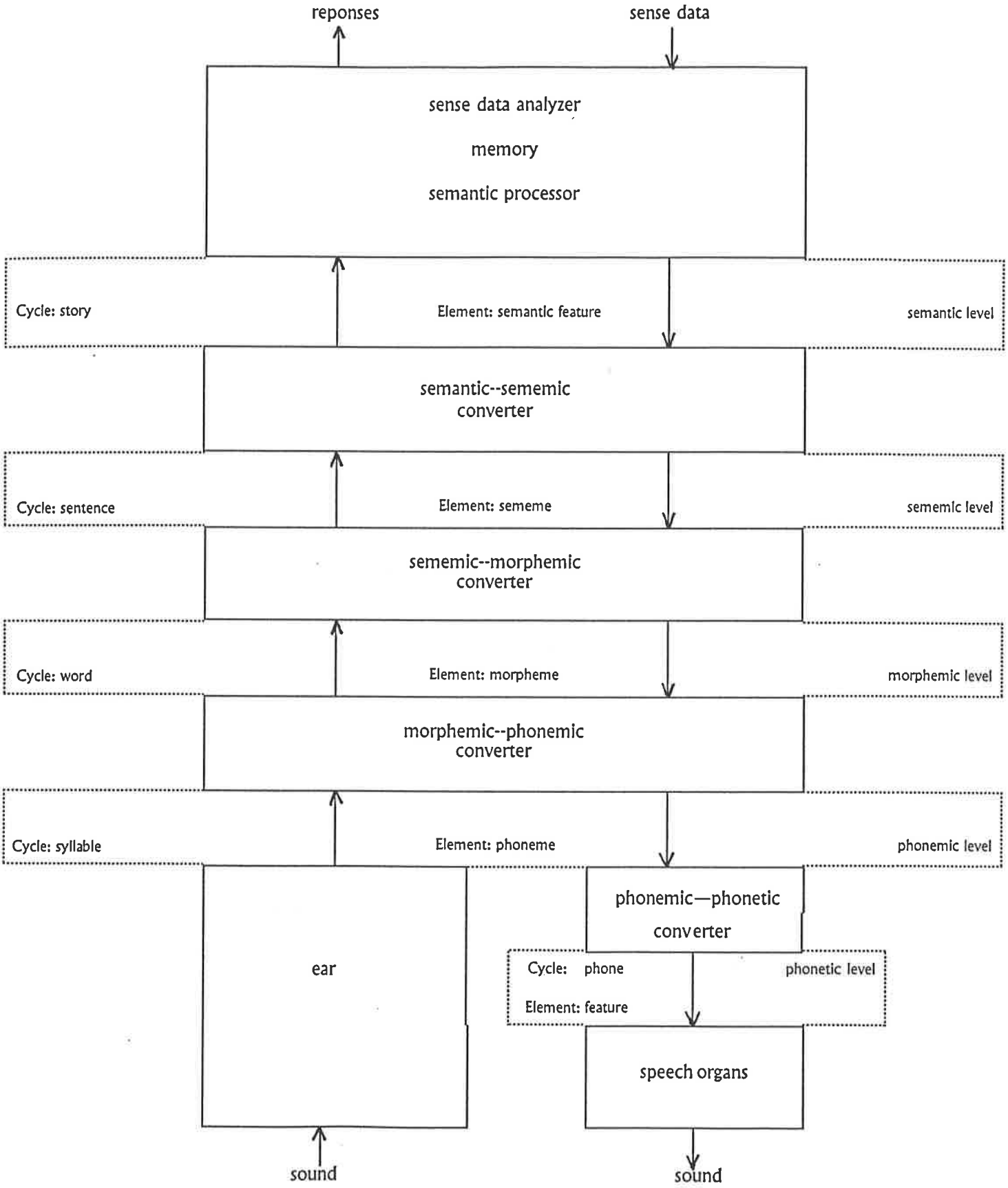


Figure 1