A Sketched Computational Theory of Language Comprehension

Wai-Kiang Yeap (yeap_wk@cs.otago.ac.nz)
Artificial Intelligence Laboratory
Department of Computer Science
University of Otago, Dunedin, New Zealand

Abstract

This paper describes a semantically based computational theory of natural language comprehension. The theory argues for a semantically rich lexicon whose entries can be described as monosemic, generative and image-like. The comprehension process uses the basic definition of a word to decide how new information is to be combined with what has been interpreted so far. Next, and more importantly, the background information is used to generate the meaning of the combined words. Other semantically based approaches are also reviewed, one each from the disciplines of AI, Cognitive Science, and Linguistics.

Background

Human language is generative in the sense that speakers can create and easily understand new sentences endlessly, despite a finite vocabulary. This generative property of language results from applying regular principles of combining vocabulary items. The question is, what is the basis of these combinatory principles? A traditional answer is that they are syntactically based (Chomsky, 1988). However, in the last several decades the alternative, that combinatorial principles are semantically based, has been explored by various researchers (Dowty, 1979; Lakoff, 1987; Langacker, 1990). Since the early 70's, AI researchers have also emphasized the use of semantics, beginning with their attempt to build wholly semantic sentence analyzers to current models which exploit both syntactic and semantic constructs (Ritchie, 1983).

Despite these efforts, no semantically based computational model has been developed which can explain adequately how language works. Most existing (AI) systems, although encoded with all kinds of knowledge, only use their knowledge as a last resort, i.e. when other mechanisms have failed. This strategy is contrary to the one which humans use, namely one which involves rapid and apparently effortless retrieval of very large amounts of 'encyclopedic' knowledge. Humans' interpretation and production of language in real time, and first-language acquisition by children suggest that such encyclopedic knowledge is immediately available to the compositional process (Hoff-Ginsberg & Shatz, 1982; Marslen-Wilson & Tyler, 1987).

Previous attempts to model the compositional process using word meanings are too concerned with what word meaning is, and what kinds of information must be encoded as sense properties. These are difficult questions to answer and perhaps should be left alone (Levin & Pinker, 1991). I have thus taken a different approach, that there is no principled way to differentiate word meaning from other knowledge (such as pragmatic and commonsense), and instead allow all kinds of knowledge to be encoded at the lexical level, at least initially (like a child learning a new word). Therefore, I ask not, what does a word mean but rather what is needed and what is made explicit at each step of the process, from input to output. In particular, I am concerned with how word meanings are combined with preceding context to create a meaning of the composite phrase and under what conditions the compositional process must be held in abeyance.

The theory will be described next before discussing related work. As pointed out above, a key feature of the theory is its emphasis on how the different kinds of knowledge about words encoded at the lexical level is used in comprehension. To capture such a possibly huge amount of knowledge, it is important to distinguish what some linguists called the basic definition of a word (Ruhl, 1989) from other related information. Hence, it is proposed that each lexical entry consists of two parts, namely, (i) a basic definition and (ii) some background information. As we shall soon see, the former provides the initial constraints for combining words and the latter provides the initial context for sense generation. The result is a rich output which I refer to as a *Mental Sketch*.

In the section on related work, the work of Pustejovsky and Boguraev (1993) on generative lexicon, Franks (1995) on sense generation and Langacker (1990) on cognitive grammar are critically reviewed. We share a common view that language is generative but we differ in how the process might be realized. The conclusion presents a summary of the main points of the theory.

A Computational Theory

Word Images

What is needed as part of a word definition in order to understand language? Consider the output produced from having parsed the phrase, *I eat*, by filling in argument structure of the verb:

Eat :who I :what ? :how ? etc.

Although who is eating is now part of the representation, one must know much more in order to claim understanding of the phrase. In particular, it is reasonable to expect the use of hands to transfer food to the mouth and depending on the food, cutlery such as chopsticks or a spoon will be used. Some kinds of food are cooked and prepared in certain ways and eating is performed with manners. The phrase, *the lion eats*, would conjure up a very different interpretation.

The lack of such contextual knowledge is a major problem

with approaches which advocate the use of an independent syntactic module prior to semantic processing. A typical solution in the past has always been to make available the context as soon as possible, thus producing systems which

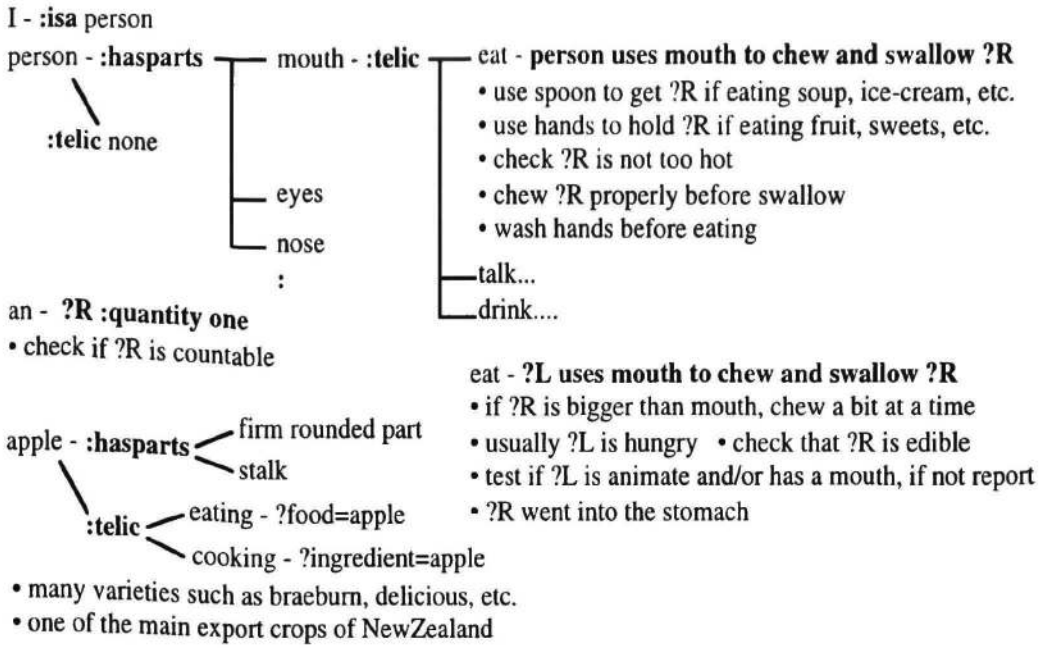

Figure 1. Examples of how some words are defined in the lexicon (see text).

process syntactic and semantic parsing in tandem. But what is context? From an implementation standpoint the real concern is how much context? If too much is allowed, the system will be slow and may not be able to process language in real-time. If too little, the information available may not be, useful enough. I would like to argue here that the question should not be how much information is made available as context but how the contextual information is to be used. This is because the amount of information depends on the situation, not on the system being given a predetermined context for each word. Similarly, such information, like the different senses of a word, should be generative.

The lexicon thus proposed here will allow as much or as little context to be encoded as part of a word definition. Naturally, some words have more and others have less but it is assumed that it is always possible to expand the context if necessary. Naturally too, one would expect that for most words, the amount of contextual information encoded is quite significance. To avoid the problem of having to process this possibly huge amount of information, a distinction is made between one's experience with the use of a particular word (i.e the context) and its basic definition. Thus, each lexical entry will consist of two parts: one part for describing its basic definition and the other for any related information. The latter is also referred to as *background information.*

Figure 1 shows some examples of how words are described in the lexicon using the above scheme. The basic definition is in bold and the background information is a list of statements and/or questions. Note that the word *I* is simplified to have the same meaning as the word *person*. Since we are not interested in implementation details here, each entry shows only what information is made explicit (and later I will argue why) but not the exact mechanism by which it should be implemented.

The :hasparts for each noun word is intended to capture its basic information. If the word describes a physical object, then visual images of the object would be its best description. For our discussion here, a list of the different parts is used instead. One also learns the function of each noun and this extra information is also included as part of its basic definition. Following Pustejovsky and Boguraev (1993) this part is referred to as its telic role. Any subpart of a noun is itself a noun and therefore, if necessary, will have its own hasparts and telic role definition, thus producing a recursive structure.

The basic definition for each verb is a process description. In it there are variable fields which need to be assigned correctly to obtain the verb meaning. Words like articles, adverbs and adjectives enhance the meaning of an adjacent word. Therefore their basic definition consists of a variable field together with some extra information. The variable field will be assigned with the definitions of the appropriate adjacent word and the extra information will then be used to enhance that definition. As we shall soon see, the extra information itself may contain variable fields which will need to be assigned as well.

Note that the syntactic category of words is not made explicit in the lexicon although the different representations mean that such information is implicitly available. Therefore, when one perceives a word and retrieves its definition, it is not its syntactic class or properties that one is attending to but a complex description of its meaning. This corresponds to the powerful imagery of words that humans so often produce on hearing a word. In this sense, each lexical entry is referred to as a word *image* and their composition would produce a complex image which is referred to as a *Mental Sketch*. Note that the two may be used interchangeably since

a Mental Sketch is but a larger image.

When combining two word images, their background information is merged. It is during this process that much of the reasoning goes on to generate the meaning of the new image. The result is that some of the background information will be highlighted (or selected) to describe the meaning of the phrase and the rest will remain as background information in the individual image. Figure 2 shows how the interpretation of the two phrases *I eat* and *Lion eats* produces different results. The important idea is that different background information is highlighted and this then provides a context for later interpretation. This contrast significantly with those approaches which identified only a subject-verb agreement but nothing else.

[I* uses mouth to chew and swallow ?food]
- if ?food is bigger than mouth, chew a bit at a time
- usually [I*] is hungry • check that ?food is edible
- ?food went into the stomach
- use spoon for ?food=soup, ice-cream.
- use hands for ?food=fruit, sweets, etc.
- check ?food is not too hot
- chew ?food properly before swallowing
- wash hands before eatng

[lion* uses mouth to chew and swallow ?food]
- [lion*] is hungry • ?food is usually a small animal
- get ?food by chasing and killing using claws and mouth
- ?food is torn to pieces and eaten raw • bones left behind

Figure 2. Interpreting the phrases, *I eat* and *Lion eats*, will produce different background information for the same verb *eat*.

It is important to stress that the background information is made explicit to indicate how the individual understands the phrase and not, as in the early work on predictive parsers, for predicting what is next. To successfully predict, one needs to know with high probability what is next and this is not possible in language (as has been demonstrated in the early work). If the background information is not used to predict what is next, how does one make use of it in the compositional process? The trick is to use the basic definition of the incoming word as the initial basis for combining words. This is described in the next section.

Computing Mental Sketches

Since it is the definition of the next word perceived which will be used to enhance the description of the Mental Sketch, it is argued that its definition should provide the (initial) basis for deciding what to do next. However, given that the background information contains mainly related information, it is argued that it is the basic definition which is most useful.

Observe that most words have images with variables as part of their basic definition (see Figure 1). This suggests that the main task in combining words is to reason how one image could be used to replace a variable in another image. Thus, when a word is perceived, its image is retrieved and checked for variable words. If found, an attempt is made to use the images in the Mental Sketch to replace those variables. If not, it will attempt to use the image itself to replace one of the variables in the Sketch (especially a ?R variable, see below). It is not always possible or necessary to add the current image to the Sketch and if this is the case, the Sketch will hold independent images (which will have to be combined at a later stage for a meaningful interpretation).

Observe also that in most languages, certain words have a preference where to look for images to replace its variables. The choice is either a preference for the image on its left or on its right. To indicate this in the representation, special variables are used (e.g. ?L for left preference and ?R for right). However, a ?R variable can still be combined with something on its left but even if it does, it is important to set up another possible parse which will check out what is on its right. If the next image turns out to be a suitable image to replace the ?R variable, then that image is preferred. In this way, one thus sets up a preference mechanism for combining words. Some example phrases indicating the need for such a mechanism are shown below:

(1) He runs fast.
(2) He runs fast food restaurants.
(3) The word 'a' is an example of an article in English.
(4) *The word a spider induces terror in some people.

Sentences (1) and (2) show the preference for adjectives over adverbial use of the word *fast*. Sentence (3) shows how a ?R variable can still be replaced by an image on its left and sentence (4) show how even for a sentence which is not grammatically correct, such preferential binding is still clearly perceived.

The above algorithm shows how one could quickly combine words without explicitly identifying syntactic categories of words and without explicitly specifying the grammar rules. However, this is only the first part of the process. The next important step is to reason whether the combination makes sense and if not, one either signals what is wrong or tries to accommodate the differences in a variety of ways (see below). To illustrate this reasoning process, consider the parsing of a simple phrase *I eat an apple*. On perceiving the first word, one simply retrieves its image, denoted by **I***, and placed in the Mental Sketch, denoted by []:

I → [I*]

The next word perceived, *eat*, has both a ?L and a ?R variable. This prompts it to look into the sketch for an image to replace its ?L variable. There is only one image in the sketch and combining the two produces the output as shown in Figure 2. Their background information is merged.

The third word perceived, *an*, has a ?R variable and this requires creating an independent image in the Mental Sketch as follows (independent images are captured as a stack):

[?R :quantity one]
[I* uses mouth to chew and swallow ?R]

Since the Mental Sketch is not empty, its images in it is checked to see whether the image of *an* can be used to replace any of the ?R variable in the Mental Sketch itself. Note that

such a replacement which is not in the preferred sequence would require that the composition be meaningful. That is, one will have to infer that the composition makes sense given the current context. For example, in the phrase *runs fast* the word *fast* is used to describe an action and *runs* is an action. This strongly suggests that it is possible to combine the two, with the definition of *fast* modifying the definition of *runs*. However, in the above example, it is not possible to combine *eat* with *an*. The fourth word perceived, *apple*, has no variable in its definition and one thus uses it to replace a variable in the Mental Sketch. This is done successfully and the result is:

[apple* :quantity one]
[I* uses mouth to chew and swallow ?R]

The process repeats itself until there is only one image on the stack or that the process cannot proceed further. The image, [apple* :quantity one] is now a completed image and once again it is used to replace a variable in the Mental Sketch. This was done successfully again and the result is:

[I* uses mouth to chew and swallow [apple* : quantity one]]
- chew a bit at a time
- usually [I*] is hungry
- one [apple*] went into the stomach
- use hands to hold [apple*]
- chew [apple*] properly before swallowing
- wash hands before eating

The presence of adjectives and prepositions in the sentence require more sophisticated reasoning when building the Mental Sketch. For example, in addition to adding extra information to another image, an adjective must first work out which aspect of that image that it is modifying> Thus, to understand the phrase a fast book, one has to find out what is "fast" about the book. This task is reflected in the basic definition of each adjective word. For example, the word fast is defined as follows: **fast - [?R :?action done quickly]**. The ?action means that the process must find out what is "fast" about ?R that one is describing.

For prepositional phrases one would utilize the basic definition of the phrase much more in order to select where it should be attached to in the Mental Sketch. For example, to parse the sentence *I saw the girl in the park with a telescope*, the phrase *with a telescope* will be attached to the verb *saw* based on the telic role of the word *telescope*. Hence when the phrase *with a telescope* is parsed, it will immediately search the Mental Sketch for a 'seeing' action and if found, attach the phrase to it. If this fails then the Mental Sketch will be searched for another possible attachment. This example also highlights the fact that although the earlier examples may suggest that interpretation takes place preferentially over very short ranges, this is only because the examples are simple ones. The issue is not short versus long range dependencies in language but how contextual information is brought to bear on the parsing process.

In summary, what is proposed is a natural language understanding process which emphasizes on its reasoning power to interpret a sentence. The lexical entry required may best be viewed as monosemic (Ruhl, 1989), generative (Franks, 1995; Pustejovsky and Boguraev, 1993) and even image-like (Langacker, 1990). Yet, another important aspect of the lexicon is that the information serves as only a guide as to how each word may be interpreted. In particular, its definition can be generalized or simplified when interpreting a particular phrase. As a last example, consider parsing the sentence *My car drinks petrol*. On combining the images of *my car* and *drinks*, the process must try to make sense of it. One possible outcome is:

[[car* :owner me] swallows ?R]
- [car*] not animate → can't drink, how?
- [car*] does not have mouth, uses what?
- check ?R is liquid
- ?R flows into the stomach?
- [car*]? is possibly thirsty

The image of the next word, *petrol*, when added to the Mental Sketch satisfies the requirement of ?R and a relation between petrol and car has to be established. Petrol flows into the petrol tank of the car, thus giving the following interpretation (⇒ indicates generalization):

[[car* :owner me] swallows petrol*]
- how? → [petrol*] pumped into the [car*]
- [car*] uses (mouth ⇒ opening to petrol tank)
- [petrol*] flows into the (stomach ⇒ petrol tank)
- [car*]? is (possibly thirsty ⇒ takes a lot of petrol)

There are many variations in language use and how well the theory can adequately explain these variations remain to be seen. A computer program is currently being implemented to test the theory and preliminary results seem encouraging.

Related Work

Generative Lexicon

Pustejovsky and Boguraev (1993) (henceforth, P&B) argue strongly for the need to have a rich and expressive vocabulary for lexical information so that it could account, among other things, for the creative use of language. Of significance is the idea of a qualia structure for each lexical entry and the use of some generative devices operating upon them to produce the required interpretations. The qualia structure proposed is a system of four relations: a constitutive role which describes the relation between an object and its constituent parts, a formal role which describes its role within a larger domain, a telic role which describes its purpose and function, and an agentive role which describes the factors involved in its origin. For example, following P&B, the qualia structure for the word car would be:

car(x)
CONST = {body,engine,...} FORMAL = physobj(x)
TELIC = drive(P,y,x) AGENTIVE = artifact(x)

An important generative device is that of type coercion

which combines words by selecting the appropriate type of information from one word so that it could be used as an argument for the other. Using the example of interpreting the phrase, *a fast car*, P&B argued that *fast* is viewed as always predicating the Telic role of a nominal and thus one would select the relation *drive* to generate the meaning "a car which can go fast". Similarly, other meanings for phrases, such as *a fast typist, a fast book*, and *to decide fast*, could also be generated. This method thus avoids the need to enumerate the different definitions made explicit in the lexicon.

Anick and Bergler (1992) outlined how the qualia structure is used to resolve metonymy and other violations of selectional restrictions when parsing complete sentences while Copestake and Briscoe (1992) argued the need to use lexical rules in conjunction with the coercion process. Although this work further supports the need to have semantically rich lexical information, they fail to question whether the qualia structure itself is adequate or not. If a richer description of a word is needed, why impose such an arbitrary structure?

Perhaps the need to impose such a structure has to do with the need for strong typing information. For example, when discussing sentences (5) - (8), Anick and Bergler (1992) are only concerned with how to describe the selectional restriction imposed by the verb *eat* on its direct object and how to accommodate *bagel* as the object of a preposition selecting for an event:

(5) John ate the bagel. (6) John ate the meal.
(7) John left after the meal. (8) John left after the bagel.

The above shows a serious lack of emphasis on the use of the "rich" semantic information made available in the qualia structure. For instance, no reasoning is afforded as to the appropriateness of attaching *eating the bagel* to the earlier part of the sentence, *John left after*. What if the sentence is *John ate after the bagel*? In this case, one needs to search for a more appropriate action, not just an event type. One might argue that such reasoning is done at a later stage, but if that is the case then one is not using much of what is available in the enriched lexicon.

There is no reason why a lot more of the information available in the lexicon should not be used to provide a context for reasoning about the sentence. As such, it is not necessary to restrict the lexicon by imposing the use of a qualia structure. Note that the significance of P&B's example in deriving a generative meaning for the word *fast* lies in the way in which the word *fast* is defined and used and not the fact that the word *car* is being defined using a qualia structure. I have provided an alternative lexicon and the accompanying process which shows how it might be done.

Sense Generation

Franks (1995) also presented a generative approach to understanding concept combination and in particular he showed how different views of privative combinations (such as fake gun, stone lion) could be computed via a two-step process. The first step combines the two words to generate its basic meaning and the second step is to generate an appropriate interpretation depending on the viewpoints needed.

Although he suggested the use of a representation which is more psychologically motivated than P&B's approach (see Figure 3), his method of generating word meanings is quite similar. For example, to understand the phrase, *stone lion*, he coerced the two descriptions by negating the central attributes of lion using what he called an MTC_R operator. The result of the coerced representation is shown in Figure 4.

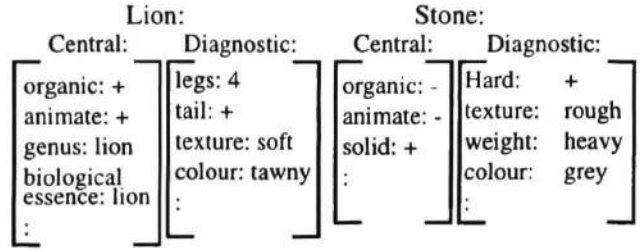

Figure 3. Each concept is represented as having both central and diagnostic features, an idea borrowed from psychological studies of humans concept formation (Medin & Shoben, 1988).

I have also used a two-level description in my word definitions (which are loosely equivalent to Frank's concept representations). However, I find the distinction, between central versus diagnostic properties unnecessary at the lexical level. Like P&B, Franks also did not utilize much of the available context in generating the appropriate sense of the combined phrase. Using the approach outlined here, the meaning of the phrase *stone lion* is obtained by first observing that both words do not have variables in their basic definition. One has to search the context for more information. If it is possible to realize that "stone is some kind of material for making ?things", then the image of *lion* would be used to replace the variable *?things*. This generates the meaning of a lion made of stone.

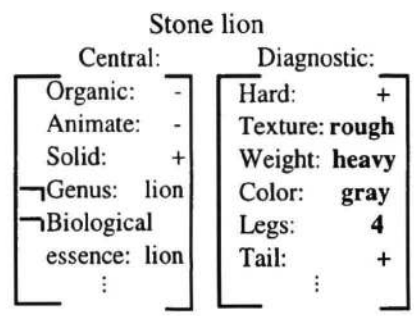

Figure 4. Interpreting the phrase *stone lion* using Franks' approach.

Cognitive Grammar

Langacker's (1990) cognitive grammar argues strongly that each linguistic expression is understood by evoking the complex conceptual descriptions of each word and combining them using a grammar which is inherently 'symbolic'. The former implies that word meanings are not described in isolation but within one or more wider domains. Each description is the product of a process of imagery, which shapes the content of a domain in a variety of ways so as to capture the ap-

propriate relationship between its salient and related features. Thus, for example, certain features of a domain may be highlighted and described at various levels of precision or at a different scale and scope. Figure 5 shows an example.

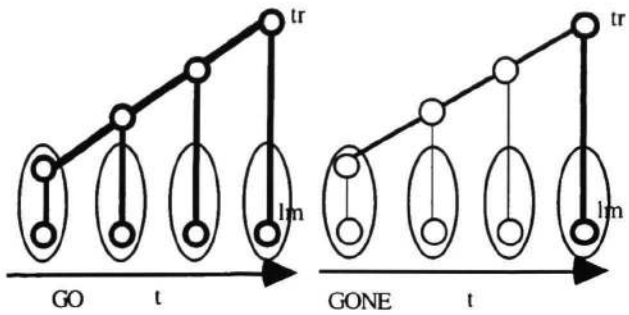

Figure 5. Different senses of *go* defined using imagery. The left shows that the word*go* is defined by showing that a trajectory (tr) is moving further and further away from a stationary landmark (lm). Using the same domain, the right shows how *gone* is defined. Reproduced from figure 4 of Langacker (1990).

The latter, claiming that grammar is symbolic, implies that grammar rules are also represented in the form of conceptual descriptions posited for representing lexical items. In general, such a description of grammar acts only as a schematic templates representing established patterns for the assembly of complex symbolic structures (see Figure 6). What is important, therefore, is that the cognitive grammar does not posit abstract deep structures from which different sentences are generated. Rather, the user has an inventory of symbolic resources of which the grammar is a part. The user must actively construct the output using these resources.

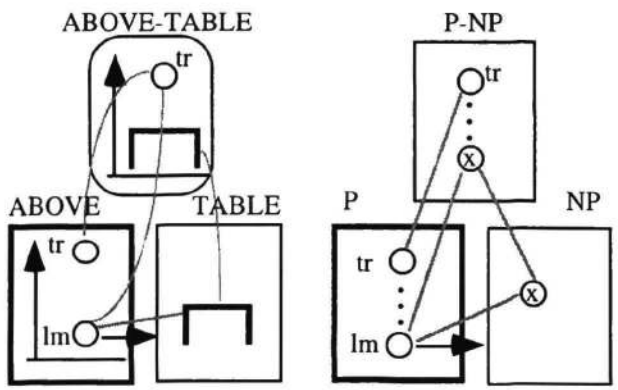

Figure 6. Representing grammar rules (right) as symbolic structure. Reproduced from figure 12 of Langacker (1990).

Langacker's idea of combining word meanings from evaluating the semantics of each word with the grammar only as a guide in the process is also central to the model proposed here. Although Langacker emphasized the complex description of each lexical item, he has not shown how the extra information is used in forming the composite structure. He did stress the importance of establishing the profile of the composite image but as shown here, it is also important to establish the necessary background information when generating the composite meaning. As for the 'imagic' characteristic of the grammar, I have shown that this is not necessary here.

Conclusion

Although the meanings of words in an expression play a crucial role in understanding that expression, it is also true that the grammar of the language gives us the order of interpretation. Thus, in the sentence *John saw Jane*, there is nothing in the semantics that will tell us who is doing the seeing and who is being seen. A semantically based approach to combining words therefore does not imply that grammar is not needed. However, it does imply that much of the effort in the process should be expended on the reasoning process based on word meanings and that much of the information should be available as part of word definitions in the lexicon.

I have shown here how this is possible and suggested a two-step process of language comprehension. The first step utilizes the basic definition of a word (which incorporates some grammatical information) and the second, its background information (which encode whatever the individual feels as important). The first step is done quickly so that much effort can be spent on the second step for reasoning about the composite meanings of the combined words. The theory is currently being tested with a computer program.

References

Chomsky, N. (1988). *Language and problems of knowledge: The Managua lectures*. Cambridge, MA: MIT Press.

Dowty, D. (1979). *Word Meaning and Montague Grammar*. Dordrecht: Reidel.

Franks, B. (1995). Sense Generation: A "Quasi-Classical" Approach to Concepts and Concept Combination. Cognitive Science 19, 441-505.

Hoff-Ginsberg, E. & Shatz, M. (1982). Linguistic input and the child's acquisition of language. *Psychological Bulletin*, 92, 3-26.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.

Langacker, R.W. (1990). *Concept , Image, and Symbol: the cognitive basis of grammar.* Cognitive linguistics research. California : Stanford University Press.

Levin, B. & Pinker, S. (1991). Special issue of *Cognition* on lexical and conceptual semantics. *Cognition*, 41, 1-7.

Marslen-Wilson, W. & Tyler, L.K. (1987). Against Modularity. *In J.L. Garfield (Ed.) Modularity in Knowledge Representation .*

Medin, D., & Shoben, E. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, 20, 158-190.

Pustejovsky, J. & Boguraev, B. (1993). Lexical knowledge representation and natural language processing. *Artificial Intelligence*, 63, 193-223.

Ritchie, G. (1983). Semantics in parsing. In *M. King (Ed.), Parsing Natural Languages*. New York: Academic Press.

Ruhl, C. (1989). On Monosemy: A study in Linguistic Semantics. Albany: State University of New York Press.