Other Things Detritus: Progress Report on
Detritus Passage Number 1
Fall Semester 2002

Jeffrey Galko

Abstract: My task has been to develop a network using SNePS that represents a given

passage of text with the unknown word 'detritus' in it. Specifically I have been

working on passage one of the set of passages containing 'detritus.' I have

concentrated my efforts into producing two different approaches to the representation

of the following sentence: "Soaked-leaf litter and other detritus yield tannin-rich tea."

My work is not only applicable to the computational aspect of the CVA project.

Perhaps a unique aspect of my work is that I have used some of the results from the

verbal protocol analysis conducted by the Department of Learning and Instruction as a

guide to represent the sentence. My work is an attempt to combine the efforts of both

the Computer Scientists and the Reading Instructors who have participated in the

project.

Most words that people have learned are learned through encountering new words in written texts. Given this, it is somewhat surprising how little is known about the strategies that readers use to learn new words. Evidence of how little is known in this area is found in Clarke and Nation (1980). They advise readers to guess the meaning of the unknown word without giving explicit guidance on how to do this. Undoubtedly learning new words involves some sort of guesswork on the part of the reader. But it involves much more. The reader must be sensitive to certain cues within the text that shed light on the meaning of the word and use his or her background knowledge and reasoning skills to reach a tentative conclusion regarding word meaning (Sternberg et al. 1983 and Sternberg 1987).

The goal of the Contextual Vocabulary Acquisition (CVA) project is to develop a better understanding with respect to the methods that people use to determine the meanings of new words. In a sense, this involves an effort to make the implicit explicit. That is to say that often times we are unaware of the mind's activities while reading. We may be employing a strategy to tackle the meaning of a new word while not even noticing it. The CVA project is an attempt to combine the

efforts of computer scientists and reading experts not only to understand the methods of readers but also to improve them through the development of a curriculum that can be taught to young readers.

Throughout the semester I have been working on passage number one in the set of passages containing the unknown word 'detritus.' There has not been work on 'detritus' prior to this semester. It was determined very early in the semester that the most critical portion of the passage was that which included the unknown word. The sentence reads as follows: "Soaked leaf litter and other detritus had yielded the usual tannin-rich tea, and the water's sleek surface was dark as buffed ebony, punctuated sparsely by large trees, their roots and buttresses submerged" (Quamman, D. (2001). End of the line: Megatransect part 3. National Geographic. August 2001, p. 75-97). The entire passage reads as follows:

> And now his path was blocked by a final obstruction, not the most daunting he'd faced but nonetheless serious: this blackwater sump, a zone of intermittently flooded forest converted to finger lake by the seasonal rains. Soaked leaf litter and other detritus had yielded the usual tannin-rich tea, and the water's sleek surface was dark as buffed ebony, punctuated sparsely by large trees, their roots and buttresses submerged. (Quamman, D. (2001). End of the line: Megatransect part 3. National Geographic. August 2001, p. 75-97.)

It was also decided before starting work that the first sentence in this compound sentence was most important, with the second providing little, if any, information

about what 'detritus' means. Accordingly, the focus of my attention was on the sentence: "Soaked-leaf litter and other detritus had yielded the usual tannin-rich tea."

It is worthy to note that in the passage mentioned above the standard physical geographic meaning of 'detritus' is used. Webster's New World Dictionary: Second Concise Edition (1982, p. 206) defines 'detritus' as "fragments of rock, etc. produced by disintegration or wearing away; debris." The other passages make use of 'detritus' in its broad meaning of debris and not in its more specific meaning within physical geography. For example, in the second passage 'detritus' is used to refer to what remains after a biological process (i.e., the leftovers of fish feeding). Also, the last passage (passage number 7) makes use of 'detritus' in the metaphorical sense of what results when energy is not invested into maintaining the organization of one's apartment.

SNePS Network #1

One way to represent the sentence "Soaked leaf litter and other detritus had yielded the usual tannin-rich tea" is shown immediately below:
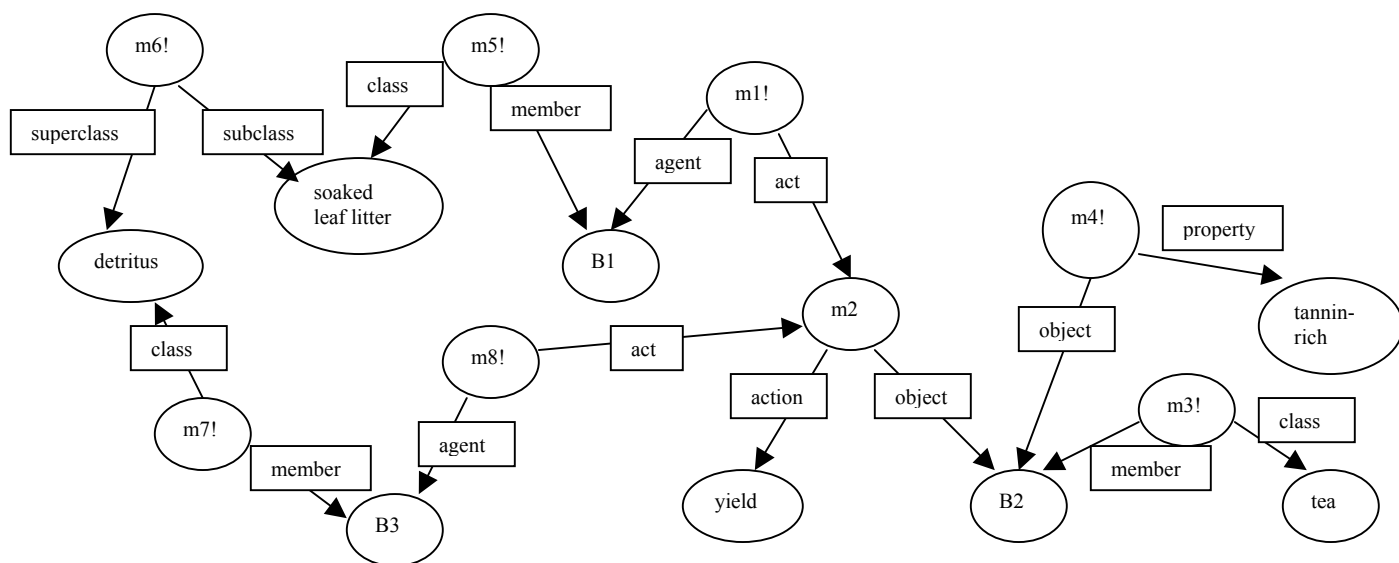
This will be referred to throughout as "the first SNePS network" or "SNePS

network #1" as it was the first network in terms of temporal order that was created.

Starting from proposition m1!, the network says the following. [[m1!]] is the proposition that some agent [[B1]] performs an action [yields] on object [[B2]]. [[m3!]] is the proposition that [[B2]] is a member of the class [tea]. [[m4!]] is the proposition that [[B2]] has the property [tannin-rich]. [[m5!]] is the proposition that [[B1]] is a member of the class [soaked leaf litter]. [[m6!]] is the proposition that [soaked leaf litter] is a subclass of [detritus]. [[m7!]] is the proposition that [[B3]] is a member of the class [detritus]. [[m8!]] is the proposition that some agent [[B3]] performs an action [yields] on object [[B2]].

One decision that must be made with respect to this sentence is how to represent the noun phrase 'other detritus'. This is done in the above representation through two steps. First, detritus is a superclass to the subclass soaked leaf litter. Second, B3 is used to express that there is something that is a member of the class detritus that yields tannin-rich tea and is not a member of the class soaked leaf litter. Of course, B3 is a base node which allows us to make assertions about individuals that meet certain descriptions without naming those individuals. This gives us a way of saying that there is an agent that is a member of the class detritus other than soaked leaf litter which we know nothing else about. The key here is to notice that the adjective 'other' is not explicitly represented in the above network.

Toward the end of the semester, the above network was built using the SNePS User Language (SNePSUL) and the knowledge base was subsequently queried. What I would like to do now is describe how the network was built. Following the

instructions of the SNePS users manual ("SNePS An Interative Approach" by Stuart C. Shapiro and William J. Rapaport) our first step is to define the arc labels. Thus, we have the following lines of code:

```
(define agent act)
(define action object)
(define object property)
(define member class)
(define subclass superclass)
```

Next we can build the network by writing the following lines of code.

(assert member #SLL class SLL)

This line corresponds to M5!. The # operator crates what is known as a Skolem constant. What this line of code says is that there is something that is soaked leaf litter, which also belongs to the class soaked leaf litter. SLL is an abbreviation for soaked leaf litter.

(assert member #tea class tea)

This line expresses m3!. #tea is B2 within the representation. This line says that there is something that is a member of the class of tea.

(assert agent *SLL act (build action yield object *tea))

This line captures propositions m1! and m2. The * operator is sued to indicate that a particular thing has done something or a particular thing has had something done to it. In other words, the soaked leaf litter yields the object tea.

(assert object *tea property tannin-rich)

This line corresponds to proposition m4!. It says that the object tea has the property of being tannin-rich.

(assert superclass detritus subclass SLL)

This line captures proposition m6!. It asserts that detritus is a superclass of soaked leaf litter.

    (assert member #detritus class detritus)

This creates proposition m7!. It asserts that there is something which is a member of the class detritus.

    (assert agent * detritus act (build action yield object *tea))

This builds proposition m8!. It says that the member of the class detritus (B3) yields tannin-rich tea.

Subsequent to building the network using SNePSUL, the knowledge base was queried. I wanted to see what the knowledge base would return when asked to describe the class detritus. The query I used was:

    (describe (find class detritus))

This produced the following return:

    (m7! (class detritus) (member B3))
    (m7!)

This is interesting because when we ask about the class detritus we obtain information about B3 which is the way we chose to represent 'other detritus.'

The next stage of the work completed this semester centered on an analysis of a verbal protocol conducted with a human subject. (The protocol was performed by K. Wieland with a human subject, M.B., on April 19, 2002.) A human subject, referred to as M.B., was given the set of passages containing the unknown word 'detritus.' The subject was then asked to "think-aloud" about how he/she arrived at a definition of the word. What I wanted to do was to perform a cursory analysis of the results of the

protocol in order to capture M.B.'s reasoning process in a series of SNePS representations.

The protocol reveals that M.B. utilizes three key known phrases: "miscellaneous stuff," "dirt or foliage," and "soaked leaf litter." It is worthy to note that the first two phrases are a part of M.B.'s background knowledge and are not explicitly mentioned in the passage. Accordingly, a key portion of M.B.'s background knowledge consists of the following categorical syllogism.
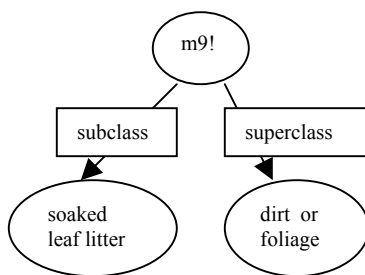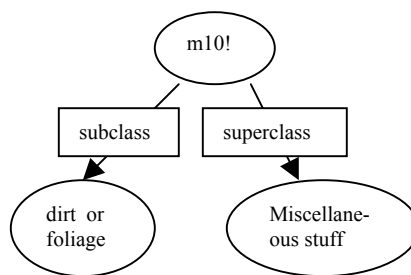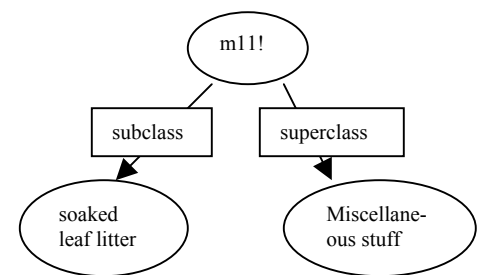

Figure 1   Figure 2   Figure 3

From the premises "soaked leaf litter is dirt or foliage" (Figure 1) and "dirt or foliage is miscellaneous stuff," (Figure 2) it is inferred that "soaked leaf litter is miscellaneous stuff" (Figure 3). The 'is' here is of course being used elliptically for "is a kind of." The conclusion of the syllogism does not need to be explicitly represented in a SNePS network as it is arrived at through path-based inference.

Using the above background knowledge combined with a series of inference rules, the subject arrives at the following two hypotheses regarding the meaning of detritus from passage number 1. The first hypothesis is that detritus is miscellaneous stuff. The second, is that detritus is dirt or foliage. These hypotheses are clearly expressed in the protocol at lines 12 and 20, respectively. The second hypothesis was elicited after

the meaning of the word 'tannin' was explained to the subject. The communication of the first hypothesis remains a bit more mysterious as the subject seems to come up with the meaning spontaneously.

SNePS Network #2

SNePS network #2 is an attempt to capture the inference rules that the subject employs to arrive at the hypotheses referred to above. For reasons to be discussed below, SNePS network #2 is not compatible with SNePS network #1. Thus, the two networks are alternative ways to represent the sentence in question.

SNePS network #2 is a set of two general rules that are expressed in the form of conditional propositions. The first is the following:
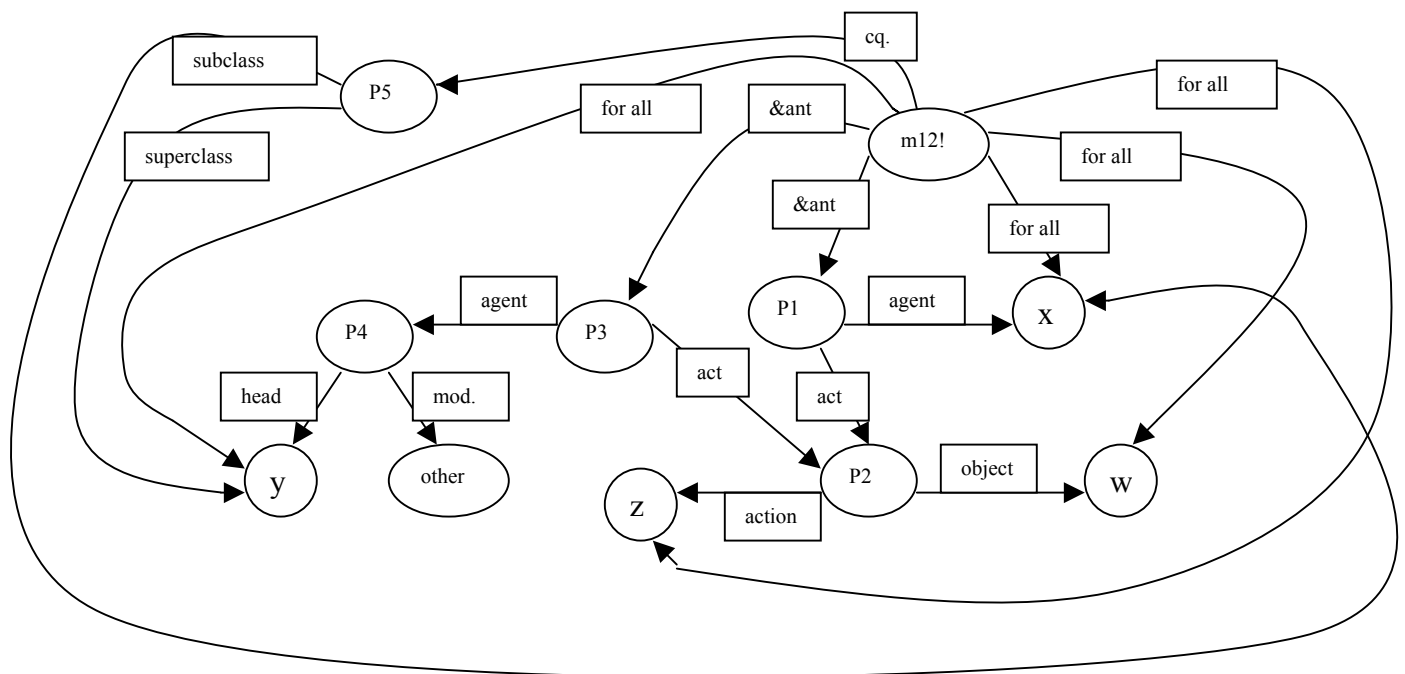
What this says in plain English is that "If x and other y do action z, then x is a kind of y." This is a very close approximation of the way M.B. seems to arrive at his/her

hypotheses. The variables x and y represent soaked leaf litter and detritus respectively. The variable z represents yielding tannin-rich tea.

The second representation combines the background knowledge of soaked leaf litter with the consequent of the first conditional to produce the hypotheses. Here is the second representation:
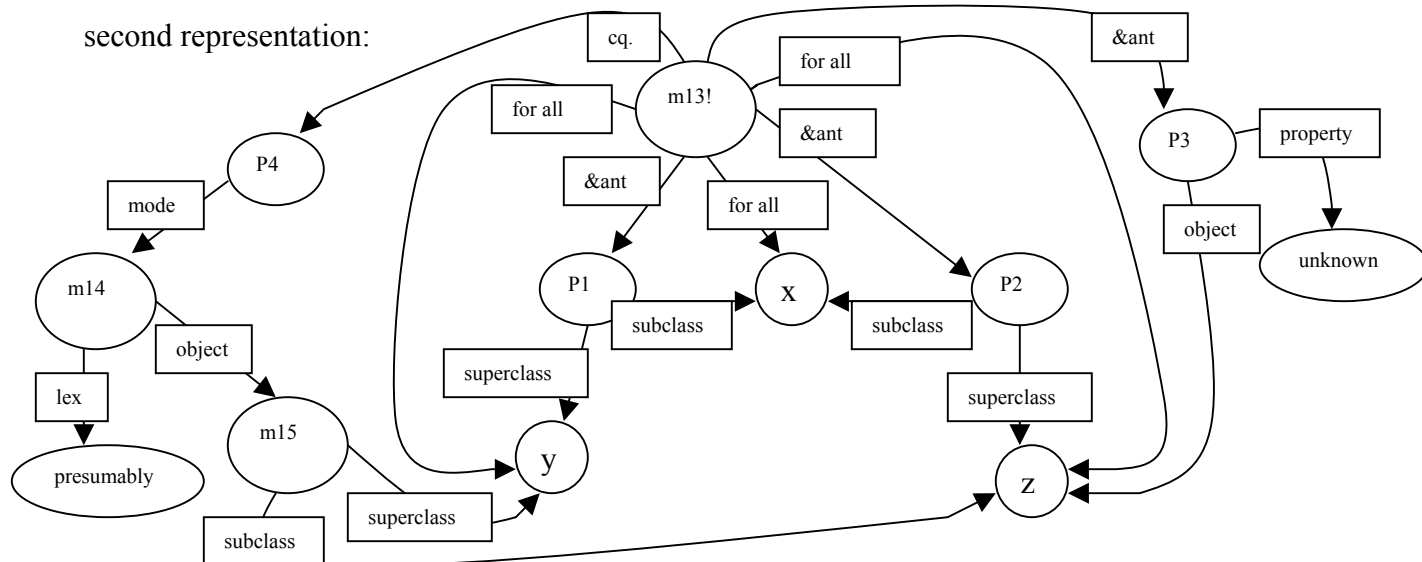
What this representation says is the following: "If x is a kind of y, and x is a kind of z, and z is unknown, then presumably z is a kind of y." In this presentation x stands for soaked leaf litter. The variable y can stand for either miscellaneous stuff or dirt or foliage. z is detritus. The first proposition in the antecedent, "x is a kind of y," is drawn from the background knowledge pertaining to soaked leaf litter. The second proposition, "x is a kind of z," is derived from the first conditional proposition shown above. The purpose of having proposition "z is unknown" in the representation is to prevent circular reasoning and unwanted inferences (e.g. y is a kind of z). In the future, it may prove unnecessary to have this proposition as part of the antecedent clause in the conditional proposition. Further research needs to be done to determine whether it is

necessary to include this in our representation. Here is what the representation would look like without it:
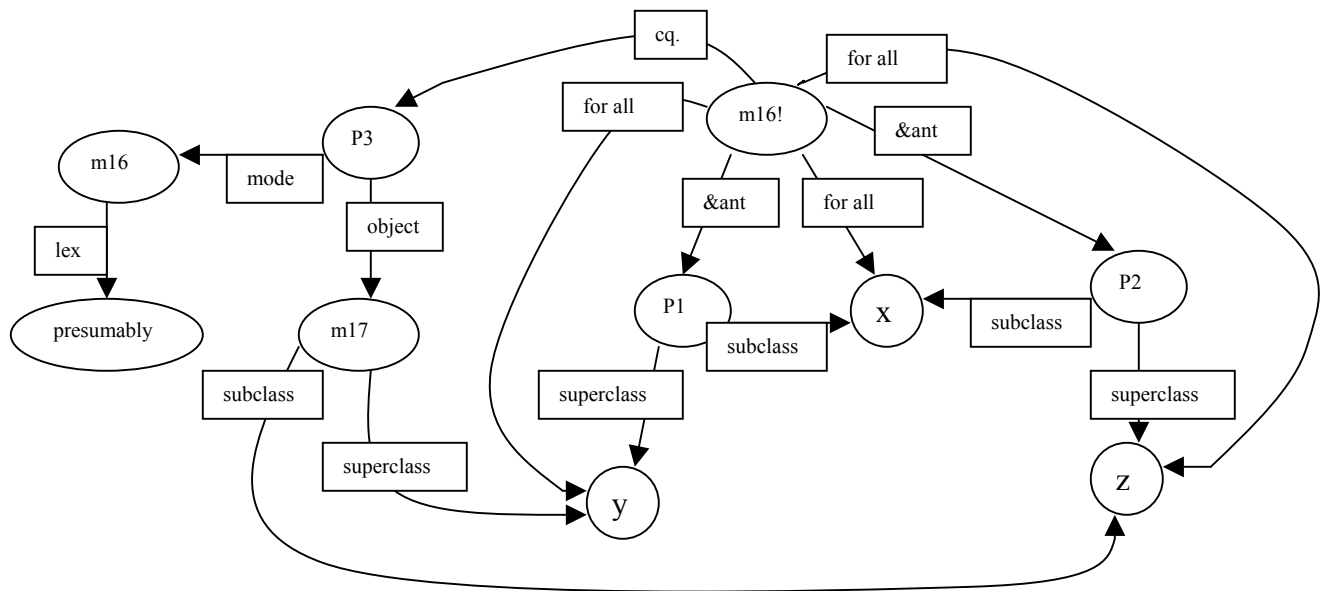
Finally, the consequent says that "presumably z is a kind of y." Presumably is used because it may in fact prove to be false that z is a kind of y. Just as the human subject is permitted to revise his/her hypotheses, our rule is defeasible. Defeasible meaning that the results of whatever reasoning that occurs can be taken back and the reasoning process undone. This would happen if further evidence was gathered that contradicted the consequent of our rule. SNePS has a belief revision system (SNeBR) built into it that allows defeasible inferences.

The Differences Between SNePS Network #1 and #2

The main difference between the two ways of representing the sentence is in the general strategies they respectively utilize. Network #2 uses a series of general rules.

These rules resemble grammatical rules in that if the word "other" appears as an adjective in a sentence, then the rule is activated. Also, if SNePS network #2 is used, it may end up producing a representation that significantly resembles SNePS network #1. On the other hand, SNePS network #1 represents the sentence "directly" without the use of general rules.

Another meaningful difference is found in the way the word "other" is represented. In network #1 other is not explicitly represented. However, in SNePS network #2, other is represented using the mod./head case frame. There is clearly no one correct way to represent this sentence. Nevertheless, the advantages of both are obvious. If we want to apply our work to other sentences of similar grammatical construction, then network #2 is the clear choice. If we want to capture the uniqueness of the sentence in question instead, then network #1 is the winner.

Future Work

The project on detritus should be extended in the following ways. First, it would be a good idea to build the SNePS network #2 using SNePSUL and compare results against SNePS network #1. Second, coding more passages including detritus would be a good idea. Also, if more verbal protocols are available in the future, they should be analyzed to determine what background knowledge subjects are utilizing to understand the word.

References

Clarke, D.F., & Nation, I.S.P. (1980), "Guessing the Meanings of Words from Context: Strategy and Techniques", *System* 8: 211-220.

Guralnik, D.B. (Ed.). (1982). *Webster's New World Dictionary: Second Concise Edition*. (New York: Simon & Schuster).

Quamman, D. (2001). End of the line: Megatransect part 3. *National Geographic*. August 2001, p. 75-97.

Shapiro, Stuart C. and Rapaport, William J. (2002). SNePS: An Interactive Approach. Retrieved from http://www.cse.buffalo.edu/~rapaport/cva.html

Sternberg, Robert J.; Powell, Janet S.; & Kaye, Daniel B. (1983), "Teaching Vocabulary-Building Skills: A Contextual Approach", in Alex Cherry Wilkinson (ed.), *Classroom Computers and Cognitive Science* (New York: Academic Press): 121-143.

Sternberg, Robert J. (1987), "Most Vocabulary is Learned from Context," in Margaret G. McKeown & Mary E. Curtis (Eds.), *The Nature of Vocabulary Acquisition* (Hillsdale, NJ: Lawrence Erlbaum Associates): 89-105.