MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS 1963 A

②

The Counselor Project

Department of Computer and Information Science

University of Massachusetts

Amherst, Mass. 01003

The Analysis of Nominal Compounds

Wendy G. Lehnert

December, 1985

CPTM #8

DTIC
ELECTE
DEC 1 6 1986
S
D
E

407701 SH

86 10 28 063

2

The Analysis of Nominal Compounds

Wendy G. Lehnert

December, 1985

CPTM #8

If one surveys the standard representational techniques for natural language that have evolved in linguistics and artificial intelligence, it is difficult to find much uniformity in the theories proposed. Distinctions are made between linguistic performance and linguistic competence, syntactic regularities and conceptual content, formal semantics and processes of common-sense inference, structural models and process models. An innocent bystander could easily come to the conclusion that the study of language is both ill-defined and lacking in systematic research methodologies. The whole business is rather reminiscent of the three blind men who conclude that an elephant is like a tree, a snake, or a wall: what you find depends a lot on where you poke around. [ἡε ᾳ..ἐ. ᾳ

It is not my intention here to sort out all the conflicting research premises and competing methodologies associated with the study of language. Instead, I will briefly identify my own position with respect to some of these larger issues, and then proceed to discuss a specific problem associated with my end of the elephant. My interest in language addresses language as a vehicle for communication. I am concerned with the conceptual content of sentences and the cognitive processes that extract conceptual content from a text or a discourse. These processes must be described in terms of human memory models and concerns for psychological validity or at least a healthy respect for psychological plausibility. But I am not a psychologist because I do not run experiments on subjects or analyze data. I conduct my research by writing computer programs that simulate language processing behavior. These complex computer programs allow me to develop theories and sometimes test competing explanations within a single theoretical framework. I am happy to borrow ideas from linguistics and psychology, but the theoretical foundations for my work come from artificial intelligence.

I have been involved with the design of many natural language systems, including question answering systems [Lehnert 1978], story understanding systems [Lehnert et al. 1983], and summarization systems [Lehnert 1982, 1984]. More recently I have been concentrating on problems specific to conceptual sentence analysis [Lehnert and Rosenberg 1985], a crucial component for virtually all other language tasks. Over the last ten years we have seen tremendous progress in this area, and more than a small amount of controversy over the relationship between syntactic and semantic analysis. Major contributions to the problem of conceptual sentence analysis have been made by a number of people in artificial intelligence (see for example, Charniak 1981, DeJong 1979, Dyer 1983, Gershman 1979, Lebowitz 1983a, Lytinen 1984, Riesbeck & Martin 1985, Riesbeck & Schank 1976, Small 1980, Waltz & Pollack 1984, Wilensky 1981).

To state the problem simply, a conceptual analyzer must input a sentence (or sequence of sentences) and return a representational structure that captures the meaning of the sentence(s). This presumes a system for meaning representation that is suitable for higher level processes of inference and memory integration. One would also like the representational structures to be canonical in the sense that paraphrases of a sentence should produce identical (or very close to identical) meaning representations. Canonical meaning representations are motivated by

2

psychological experiments which show that people are able to retain the general meaning of a sentence long after they have forgotten the precise wording of that sentence [Bransford & Franks 1971].

People who build conceptual sentence analyzers are necessarily concerned with the representations their analyzers produce, and all of the various arguments associated with competing representations. Syntactic parse trees are routinely dismissed since they were never intended to capture meaning in the first place. Lexical predicates and case frames do constitute an attempt at meaning representation but they tend to fall short of canonical form criteria because the predicates are typically determined by the main verb. For example, the two sentences

"John gave Mary a book."
"John sent Mary a book."

result in two distinct predicate representations:

give (John, book, Mary)
send (John, book, Mary)

Extra rules have to be invoked if we are going to understand that these two representations are conceptually close. There is also a problem with lexical representations in that they do not address the problem of word sense ambiguity. The main event underlying "John gave Mary a book" cannot be distinguished from the events described by, "John gave Mary a kiss," or "John gave Mary a disease." Lexical representations do not attempt to make any such distinctions. To develop a representation that can capture appropriate word sense distinctions we have to resort to "deep" semantic representations which require decomposition into primitives.

A few people have developed conceptual representations based on primitive decomposition [Wilks 1972, Schank 1975, Lehnert 1979], and many systems employ primitives as a practical matter-of-course without a strong theoretical basis. Suffice to say, there is no "standard theory" for conceptual representation and it is likely to be a while before anyone presumes to offer a comprehensive solution. Many researchers believe we are pursuing a form of "applied epistemology" in which case it is not feasible to expect elegant solutions of great generality. The knowledge required to understand a newspaper story is different from the knowledge required to read a technical paper on hydrodynamics, so there is every reason to expect that the internal meaning representations associated with such disparate tasks will themselves be altogether different. Each knowledge domain may require its own set of event primitives and representational devices. Much exploratory work remains to be done before we can hope to look for general patterns and universal techniques.

Nominal Compounds

It is within this exploratory spirit that I would like to focus on a specific natural language construct that relates strongly to the question of internal meaning representations: nominal compounds. Simply defined, a nominal compound is a noun phrase consisting of nothing but nouns:

> dog leash
> computer terminal
> car mechanic
> leather book jacket
> steel doorknob plate screws

Within linguistics, nominal compounds have been studied and discussed for a number of languages. The major studies for English include [Koziol 1937, Jespersen 1942, Lees 1960, Marchand 1969, and Levi 1978]. Nominal compounds have a shorter history within artificial intelligence but there has been some work addressing the phenomenon [Russell 1972, Finin 1982, Gershman 1979, Lebowitz 1984a]. From the perspective of meaning representations, nomimal compounds pose an interesting problem for information processing models. People manage to make sense of these noun phrases with little or no conscious effort. But the conceptual content of a nominal compound presumes a subtle understanding of semantic features and implicit relationships between concepts. There are structural regularities within nominal compounds that suggest a form of syntax, yet no one is ever taught these rules in school.

For example, no one thinks twice about the meanings for:

> a dog collar vs. a flea collar

Yet the natural conceptual interpretations associated with these phrases are clearly distinct. A dog collar is a collar worn by a dog, but a flea collar is a collar worn by a dog or cat for the purpose of killing fleas. Conceptual distinctions can be found in any number of otherwise simple examples:

oil truck	(a truck that carries oil)
diesel truck	(a truck that runs on diesel fuel)
paper filter	(a filter made of paper)
coffee filter	(a filter for filtering coffee)
dress watch	(a watch to wear with formal dress)
gold watch	(a watch made of gold)

toy car	(a car that is a toy)
business car	(a car used for business travel)

In some cases, entirely different word senses must be selected for the head noun:

play pen	(an enclosure)
fountain pen	(a writing instrument)

butcher block	(a piece of wood)
city block	(a region)

salary cap	(a maximal amount)
wool cap	(a hat)

Other features of nominal compounds surface when we examine longer ones:

plastic cat food can cover
(a cover made of plastic that goes on a can filled with food for cats)

There are rules that regulate the proper word order for a noun group of this complexity. We cannot refer to a "can food cat plastic cover" or a "plastic food cover cat can," and expect to make any sense.

These examples suggest a number of interesting problems associated with nominal compounds which we will discuss from a representational perspective. The problems we will consider are not well understood, but they may provide us with some intriguing glimpses of conceptual information processing at work.

Lexicalization

One representational strategy for nominal compounds is commonly referred to as lexicalization. A noun phrase is considered to be "lexicalized" if it assumes the status of a lexical dictionary entry. In this case the noun phrase effectively becomes a single vocabulary item that happens to contain some blank spaces thrown in among the alphabetical characters. In artificial intelligence, this same idea has been suggested by the notion of a "phrasal lexicon" [Becker 1975].

In a computational model, a lexicalized noun phrase can be handled very simply. All we have to do is make sure we can find the lexicalized phrase in our dictionary, and then we can retrieve any ready-made definition we wish to associate with the given phrase. There is no need for "run-time" analysis or complicated procedures that attempt to construct a meaning representation dynamically. It is all just a straightforward problem of dictionary access. The idea of limited

lexicalization is quite reasonable. Certain phrases do seem to be rather specialized in their meaning, and it is unlikely that anyone would ever manage to construct their meaning from general principles. Some likely candidates for lexicalization include:

night stand   stop watch
hay ride   water closet
ball park   bank shot

It seems quite plausible that these phrases are simply memorized as units and effectively treated as lexical items. It is not possible to paraphrase these expressions as we might with other nominal compounds. A "dried grass trip" just doesn't work as an alternative to a hay ride and "sphere park" is not likely to conjure images of hotdogs and crowds. Further evidence for lexicalization can be found in the etymology of agglutinated words. Many words start out as noun groups and then attain full lexical status through popular usage:

sundial   gunrunner
hellfire   shoelace
motorboat   mothball

But the extreme position on lexicalization claims that all nominal compounds should be handled as pre-defined dictionary entries. This position is not feasible for computational models of language processing since it requires that we anticipate all possible noun groups. In principle, one could argue that this presupposes an infinite dictionary, since it is possible to construct noun groups of arbitrary length. An argument based on infinite dictionaries is admittedly stretching things a bit since there is a limit to how much people can actually handle in practice. It is possible in some sense to use the phrase:

"plastic cat food can cover retail package manufacturer rebate deadline"

in referring to the deadline on a rebate offered by a company who packages for sale the plastic covers that go on cat food cans! But one would not expect to find such a construct in reasonable discourse. So we will dismiss the infinite dictionary argument as an argument about linguistic competence.

From a performance perspective, we can still combat the extreme lexicalization position by noting that strong regularities govern compound noun groups. We must assume that regularities have evolved in language because they make it easier for people to process language quickly and effectively. If all noun groups are handled by a simple process of dictionary access, what purpose is served by their structural regularities? How is it possible for people to make sense of a new noun group the first time it is encountered? In time, a specific noun phrase that is frequently used may become lexicalized for a given speaker. But that does not preclude the

possibility of dynamic noun phrase analysis for other speakers, nor does it follow that all lexicalized noun phrases start out as such. As one becomes more familiar with the concept underlying a frequently used noun phrase, it may be that lexicalization occurs as a side-effect of concept acquisition. The relationship between lexicalization, conceptual analysis, and concept acquisition appears to be a worthy problem for the cognitive scientists among us.

In any case, we will assume that people have some cognitive facility for processing novel noun groups and extracting conceptual content from such constructs without the benefit of a phrasal lexicon. In order to better understand what this facility must accomplish, we must examine the regularities present in nominal compounds and consider the underlying memory mechanisms that exploit these regularities.

Specialized Meanings

If we hope to find general mechanisms for meaning extraction, we must understand that such mechanisms will be limited and may not be capable of producing certain associations commonly attributed to complex nominals. Many of these idiosyncratic associations can be attributed to specialized domain knowledge and are often described as extralinguistic because they depend on a speaker's education, ethnic background, occupation, and so forth.

For example, about ten or twelve years ago the term "fern bar" surfaced in American discourse. The term was popular among urban singles who frequented the bar scene. It referred to bars that catered to a particular type of upwardly-mobile clientele (forerunners of the YUPPIE phenomenon). Many of these bars shared a common decor: exposed brick walls, the menu on a blackboard, oversized plate-glass windows, blond woodwork, and lots of hanging ferns. People began to refer to these establishments as fern bars. The reference was mildly derogatory, as it conveyed a hint of condescension toward those misguided souls who would tolerate overpriced drinks for the sake of running with the in-crowd. So on one level, the term meant a bar containing ferns, but it also carried a specialized meaning that went far beyond that simple concept. By now, the term is so specialized that it is possible to have a fern bar without any ferns.

Whenever a term carries a specialized meaning which cannot be derived by general mechanisms, that term must be lexicalized (for those speakers who understand its extralinguistic associations). If it weren't, how would those associations be accessed? Many complex nominals, especially noun-pairs, carry specialized meanings that cannot be derived by general mechanisms. The question of how specialized meanings are acquired is an interesting one. I can recall the first time I heard the term "fern bar" in casual conversation: I understood immediately what was meant by

the reference. The concept had been present in my mind beforehand, but I had no concise linguistic description for easy reference. The name fit the concept and the specialized meaning required no explanation for me. It is probably more often the case that specialized meanings have to be described and explained.

Whatever the process of concept acquisition is for specialized meanings, we will not attempt to account for that aspect of conceptual understanding with general mechanisms of meaning extraction. Specialized meanings require lexicalized dictionary entries, so there is no point in trying to account for their presence on the basis of conceptual analysis. We will be satisfied to acknowledge their existence, and then concentrate on those constructs that can be processed by general processes of conceptual analysis.

Evidence for Conceptual Analysis

The processing of a noun phrase can be complicated by the fact that there is sometimes uncertainty as to exactly when a noun phrase is being terminated. Under certain circumstances, these situations result in garden pathentry sentence processing. A garden path sentence is one that leads the understander "down the garden path" into an incorrect interpretation as the sentence is being processed. Once the error is caught, it is then necessary to revise the original interpretation by going back and trying to understand the sentence one more time. Garden path sentences can result from syntactic ambiguities as well as lexical ambiguities, but we will restrict this discussion to those that occur because it is unclear whether a particular word should be interpreted as a noun or a verb. Such ambiguities are easy to construct:

sentry stands	paper wraps	log rolls
building blocks	dog runs	wood tops
silk dresses	baby toys	steel slides

Various mechanisms have been proposed to account for the ways that people process potentially ambiguous word pairs of this type. For example, one could assume that there is natural preference for noun-noun interpretations wherever possible, so noun-noun interpretations are always tried first. But if this were true, garden path sentences could only result from constructs that should be taken as noun-verb combinations. In such cases a noun-noun interpretation would cause a garden path before the correct noun-verb interpretation could be considered. But there seem to be many examples of garden path sentences that start out as noun-verb interpretations and end up as noun-noun interpretations:

The sentry stands were rusty.	(vs. The sentry stands alone.)
The toy rocks are green.	(vs. The toy rocks back and forth.)
The knife cuts hurt.	(vs. The knife cuts poorly.)

8

If you were not conscious of any garden path processing when you read the above sentences, you might like the idea of a limited look-ahead capability [Marcus 1980]. When a sentence is processed with limited look-ahead, one is effectively allowed to take a peek at some fixed number of subsequent sentence constituents (or words) before committing the system to any particular interpretation. By using limited look-ahead, it is possible to argue for deterministic sentence analysis, which means that it is never necessary to revise an interpretation (unless the critical information needed is further away than the look-ahead facility can go). But limited look-ahead theories have difficulty making predictions about which sentences will cause people to follow a garden path.

Given any fixed length for look-ahead, it is possible to construct a sentence with a local ambiguity that is not resolved until after the look-ahead capability is exhausted. But sentences can be found that confound even the most minimal look-ahead requirements. For example, people do not have any difficulty understanding a sentence such as:

"The prime number 2 is the only even prime."

Yet many people have difficulty understanding the following as a complete and grammatical sentence:

"The prime number few."

Both sentences are potentially ambiguous up to, "The prime number ..." So why does only one sentence cause difficulty? In these cases, the limited look-ahead model is not feasible, since the look ahead needed for "The prime number few," is as minimal as possible (one word). If limited-look-ahead is what people use, both of these sentences should be understood without garden paths.

Of course, one cannot pose serious arguments about psychological validity on the basis of subjective judgements. In this case, it is useful to collect reaction time data in order to test one's process model. If a sentence is processed as a garden path sentence, it will take longer for subjects to understand the sentence than would otherwise be the case. Since any potential garden path can (in principle) be understood without going down the garden path, it is possible for some subjects to understand such a sentence only after considering the garden path while others understand the sentence the first time through without any difficulties. If the resolution of a local ambiguity is made randomly, we would expect to see both types of processing: roughly half of a subject pool should go down the garden path while the remaining subjects do not.

A purely syntactic parser like the one proposed by Marcus [op. cit.] must be arbitrary when it cannot resolve an ambiguity by looking ahead. But reaction time data indicates that people are not arbitrary when they confront potential garden path sentences [Coker & Crain 1979, Crain & Coker 1979]. Rather, there appear to be semantic processes that influence the resolution of local ambiguities. In one set of experiments, local ambiguities of the noun-noun vs. noun-verb type were studied [Milne 1982]. After controlling for variables in syntactic complexity, it was found that the reaction times collected for potential garden path sentences could be explained by the "Semantic Checking Hypothesis." This model argues for a semantic process which is triggered whenever a word pair is encountered that could be interpreted as noun-noun. This process examines the word pair and returns some judgement as to whether or not the word pair should be interpreted as a noun phrase. If the judgement favors a noun phrase interpretation, that is how the ambiguity is resolved. If the judgement for a noun phrase is weak, alternative interpretations are then considered before resuming sentence analysis. So the "Semantic Checking Hypothesis" considers noun-noun interpretations first, but may reject the noun-noun interpretation on purely semantic grounds before continuing with the sentence.

As Milne admits, his experiments have not attempted to take into consideration possible predictive factors from a larger sentence processing context. A sentence that requires garden path processing in isolation may not be a garden path sentence when it is encountered in some larger context. But his experiments do argue convincingly for a semantic mechanism that must play at least a part in the overall process model.

We cannot say precisely how Milne's semantic checking mechanism operates. Does it compute a confidence rating on each potential noun phrase and then accept the noun-noun interpretation only if that confidence passes threshold? Or is the judgement purely binary (a yes or a no)? More importantly, how is the conceptual content of a noun-noun construct represented in memory? If we find a likely noun phrase interpretation, how do we encode that interpretation in memory? One can argue for surface representations in the case of certain lexicalized noun pairs, but other noun pairs must be encoded at a deeper level. If you ask someone to paraphrase a newspaper story about an "alleged killer," it is quite likely that the paraphrase might refer to a "murder suspect" or even "murderer. " Similarly, a "highrise office complex" might become a "skyscraper", and a "cellar door" could become the "basement entry."

The representational issues associated with nominal compounds are sometimes subtle but nevertheless intriguing. Complex noun groups may provide us with a useful "window on the mind" if we approach them with care.

Recoverably Deletable Predicates

Within linguistics, nominal compounds can be studied in terms of grammatical derivations. Some linguists have claimed that the semantic relationships within complex nominals are impossible to categorize, leaving the the problem of systematic derivation far beyond the machinery of generative grammar [Jespersen 1942]. But recent efforts suggest that this need not be the case. In particular, Judith Levi has suggested that there one type of derivation that accounts for most complex nominal surface structures [Levi 1978].

According to Levi, the majority of semantic relationships can be characterized as recoverably deletable predicates (RDP's). More importantly, it is possible to break this class of predicates into nine subgroups. We will list these nine groups here along with some examples:

CAUSE	tear gas	drug death	cigarette smoke
HAVE	picture book	lemon peel	church bells
MAKE	daisy chain	glass jar	coffee machine
USE	steam iron	water pipe	charge card
BE	target site	passport book	soldier ant
IN	house dog	kitchen table	spring breeze
FOR	dog collar	baby food	student government
FROM	sea breeze	olive oil	country crafts
ABOUT	tax law	price war	product brochure

According to Levi, these predicates have been deleted by a transformation that changes relative clauses into complex nominals. For example, "student government" is derived from "government which is for students" and "water pipe" comes from "pipe which uses water."

At first glance, these nine categories seem reasonable enough, but if one tries to categorize random noun pairs using this taxonomy, it is sometimes difficult to know which category to use. Should "cigarette smoke" be "smoke which is caused by a cigarette" or "smoke which is made by a cigarette"? Is a "dog collar" a "collar which is for a dog" or a "collar which a dog has"? While some noun pairs seem to qualify for more than one deletable predicate, many semantic relationships fail to be differentiated in this taxonomy. A "kitchen table" describes a locative relation, while a "spring breeze" describes a temporal relation. But both kitchen tables and spring breezes fall into the single predicate "in." For that matter, one could distinguish a number of semantic relationships that are only generally represented by the "in" predicate:

INHABIT	house dog	(a dog which inhabits a house)
GROW-IN	water lilies	(lilies which grow in water)

DURING	spring breeze	(a breeze which occurs during spring)
LOCATION	kitchen table	(a table which is located in the kitchen)
PART-OF	car lighter	(a lighter which is part of a car)
ABSTRACT-LOC	system bugs	(bugs which are associated with a system)

If one wishes to work with a set of RDP's, what criteria should be used in selecting the proper set? How much generality or specificity is needed? According to Levi, an RDP analysis must capture "... all the *productive* aspects of complex nominal formation and thus, all the information that a grammar can in any case be expected to predict." That is, Levi is interested in grammatical derivations which can be characterized by semantically equivalent paraphrases. It is enough to know that each noun pair associated with a given RDP can be paraphrased by the same relative clause construction:

house dog	(a dog which is in a house)
water lilies	(lilies which are in water)
spring breeze	(a breeze which occurs in the spring)
kitchen table	(a table which is in a kitchen)
car lighter	(a lighter which is in a car)
system bugs	(bugs which are in a system)

It is not necessary to break apart the "in" RDP any further since each of these different semantic relations can be expressed by the same relative clause using the preposition "in." If the relative clauses are semantically ambiguous, the associated RDP should remain ambiguous as well. As a model in generative linguistics, the notion of a recoverable deletable predicate may provide an adequate description of compound nominals. Unfortunately, RDP's are not adequate for computational models of language processing. In order to construct the semantic checking mechanism proposed by Milne, we must build a memory model that allows us to assess the credibility of concepts underlying noun groups. Any such credibility measure must be based on fine-grained semantics: if our language allows us to make periphrastic distinctions between two instances of an RDP, these distinctions must be grounded in the underlying semantic component. Even a conservative view of semantic representation would claim that our semantic foundation must be *at least* as refined as our linguistic expressions. Of course, one does want to limit the link-types allowed in a semantic network, and the selection of appropriate links is no small task. What would happen if we tried to build a memory model that was too "coarse?" Let's see what would happen if we had a semantic memory that used only links based on Levi's RDP's.

Consider the concept node for a "truck." A truck has parts: an engine, tires, doors, etc. So we would have HAVE links going from the truck node to all the various truck parts. Trucks also run on oil and either gasoline or diesel fuel so we would have USE links pointing to oil, gasoline, and diesel fuel. Trucks are also

used for carrying things, so we could have FOR links going to all the things that trucks carry: cement, lumber, freight, etc. Now consider the problem of understanding two noun pairs:

oil truck

diesel truck

In principle, these are both ambiguous. An oil (diesel) truck could either be a truck that carries oil (diesel) or a truck that runs on oil (diesel). But most people would be inclined to agree that an oil truck is a truck which carries oil while a diesel truck is a truck which uses diesel fuel. One could explain this by claiming that these terms are lexicalized, but that would be pushing lexicalization further than is necessary. It is much easier to account for these interpretations in terms of a memory model that is somewhat stronger than the one we've described.

Everyone knows that *all* trucks burn oil, so it would be uselessly redundant to describe a truck as an oil truck if that description only referred to the USE link between trucks and oil. On the other hand, it is not the case that all trucks run on diesel fuel, so a diesel truck should be interpreted in terms of the USE link much more readily. This distinction cannot be made on the basis of a simple USE link. It must rely on more refined memory associations, such as NECESSARY-USE and OPTIONAL-USE. One could further argue that the optional links pointing to gasoline and diesel fuel should themselves be conjoined by an OR link, but we cannot make that argument on the basis of compound nominals alone. If we are willing to look beyond complex nominals just a little bit, we can see how semantic interpretations based on Levi's RDP's would not be adequate for many simple analysis tasks. Consider the sentence:

Boomer's dog collar was too tight.

Anyone who understands this sentence will conclude that Boomer is a dog. Now consider:

Boomer's flea collar was too tight.

Here we may feel uncertain about whether Boomer is a dog or a cat, but we are not likely to consider the possibility that Boomer might be a flea. Unfortunately, this is the only possibility that a simple FOR predicate could produce. If a "dog collar" is a collar which is for dogs, and a "flea collar" is a collar which is for fleas, then there is nothing we can use to distinguish the conceptual meanings underlying these two senses of "for." The missing concepts are crucial:

dog collar = a collar which is worn by dogs

flea collar = a collar which is worn by dogs or cats in order to kill fleas.

It is possible to produce lots of examples like these. If our goal is to produce conceptual representations for sentences, we are going to have to go beyond the simple notion of RDP's in designing our underlying memory representation.

Semantic Categories

If we want to move beyond the notion of general surface predicates, it will be necessary to organize memory according to some set of semantic categories. Originally, semantic markers were used to resolve word sense ambiguities [Katz and Fodor 1964]. In order to analyze the conceptual content of noun groups, we will use semantic markers to resolve conceptual ambiguities that explain how words relate to one another. In creating a system of semantic memory, we are confronted with the usual problem of choosing appropriate semantic categories. The arguments that can be made for a given set of categories are ultimately empirical, and therefore critically dependent on actual computer programs that can be run to conduct experiments. For this reason, it is instructive to look at an early computer program for noun-pair processing that created multiple representations for ambiguous noun-pairs, and ranked these competing interpretations according to their credibility [Russell 1972].

In a program designed by Sylvia Russell, the analysis of noun-pairs was based on a set of 21 possible dependency functions between two nouns. Each dependency function specified semantic criteria for its application, and a resulting conceptual representation consistent with the theory of conceptual dependency [Schank 1975]. If more than one dependency function could be applied, multiple interpretations were ranked according to a fixed priority assigned to each dependency function. It was therefore possible to compute only the "most likely" interpretation by applying each dependency function in the order of descending priorities until one function was found to apply. A few dependency functions are described below:

FUNCTION NAME	PRIORITY	EXAMPLE	CRITERIA
COMPOSITION	20	rubber knife	N1 = +matter and -fluid N2 = +phys and -animate
R_GOAL	17	kitchen table	N1 = +env and +man-made N2 = -matter and +man-made
POSSESS	10	horse shoe	N1 = +animate or +inst N2 = +man-made (use: read, wear)

PART_INAL_RECIP	2	arm chair	N1 = +part

N2 = obj or category of objs of
which N1 is an inalienable part

In examining the semantic criteria required by each dependency function, we see that this model of noun-pair processing does require some semantic structures that go beyond strict semantic features. For example, in order for PART_INAL_RECIP to apply, we must be able to confirm the N1 is in fact a part of N2. Such semantic checks cannot be handled by semantic features alone. Similarly, POSSESS requires that N2 refer to an object that can be read or worn, and more importantly, read or worn by the referent of N1. A "horse tie" should not be interpreted the same way as a "horse shoe."

Russell designed this system to fit into the framework of conceptual information processing being developed at the time by Roger Schank [op. cit.], so there was no reason for her to restrict her dependency criteria to semantic features alone. However, some of the tests required by her system could not have been implemented with full generality, at least not on the basis of the semantic memory model described by Russell. For example, Russell never tells us exactly how one is supposed to know that a horse can wear shoes but not a tie. A memory model that does not distinguish between horse shoes and human shoes requires some more work.

However, Russell's work is provocative and well worth exploring. What kind of memory access is required beyond the semantic features? Do the features suggested by Russell really provide reliable coverage for a large class of examples? Do the 21 dependency functions provide broad coverage? Are static function priorities adequate or should credibility judgements be sensitive to the relative strength with which the function criteria are satisfied? What would it mean to satisfy a dependency criterion with relative degrees of strength?

Whatever the answers are, at least Russell attempted to produce a conceptual representation for the noun-pairs being analyzed. It seems that Russell was working under something of a disadvantage in that regard, since the formalisms of conceptual dependency are strongly event-oriented. If it were possible to map her concepts into a system of primitive decomposition that were more object-oriented, the resulting interpretations might be even more impressive. One such system has been proposed [Lehnert 1979], but its application to nominal compounds has never been investigated.

Semantic Memory vs. Episodic Memory

In describing her work on noun-pair analysis, Russell was careful to characterize her enterprise in terms of certain limitations. She states, "... all our features represent conceptual rather than cultural knowledge. The features generally

satisfy the criterion of being "inherent" properties rather than unstable situations or conditions." The distinction she makes can be properly described as the difference between semantic memory and episodic memory. This distinction was first popularized by Endel Tulving [Tulving 1972] as the difference between universal knowledge and knowledge about specific individuals or events. For example, semantic memory would be responsible for telling us that cats are animals, but episodic memory would tell us that Morris is a cat who sells cat food on television commercials. For many years, it was assumed that only semantic memory was required for sentence comprehension. This assumption seemed reasonable at the time, but only because no one was looking beyond single isolated sentences. As researchers began to consider the inference processes required for connected text (paragraphs or stories), a case for the importance of episodic memory in language processing begain to emerge [Schank and Abelson 1977].

In order to make appropriate inferences that connect sentences together, it is necessary to access complex knowledge structures. These knowledge structures can provide information that is only implicitly present in the original sentences, thereby allowing the reader to complete important causal connections and fill in many details that were never explicitly mentioned. If a knowledge structure is well-designed, it organizes information efficiently so that irrelevant information does not need to be accessed or examined. One famous example of a scriptal knowledge structure was popularized by Roger Schank [op. cit.] in order to explain how people make inferences about stereotypic behavior in restaurants.

While episodic memory structures are enjoying a healthy following within natural language processing circles, it is very difficult to say where semantic memory stops and episodic memory begins. Semantic markers are presumably a feature of semantic memory, but everything else has become somewhat controversial. Consider a typical proposition about penguins:

Penguins have skin.

This certainly looks like a universal truth, which should qualify it as semantic knowledge according to Tulving. It is easy to construct an "is-a hierarchy" to handle the appropriate generalizations we would like to manage. For example, we would probably not want to store the feature skin under penguins directly. It is better to create a structure which can tell us that a penguin is a bird and birds have skin. It is then up to some deductive retrieval mechanism to figure out whether or not penguins have skin. The same general mechanism would be used for propositions about canaries, chickens, and so forth and so on. But is this how people store such information? An episodic alternative is also possible. Suppose you are asked the question "Do chickens have skin?" and you are someone who loves fried chicken. If you eat a lot of fried chicken, and take particular pleasure in eating the crisp skin of fried chicken, you might access episodic structures about

eating chicken in order to answer that question. Figure 1 shows a fanciful picture of semantic penguin memory vs. episodic chicken memory to illustrate the difference more concretely. The general point is simple: given any "semantic" proposition, it is possible to imagine an episodic structure that could encode the information under consideration.
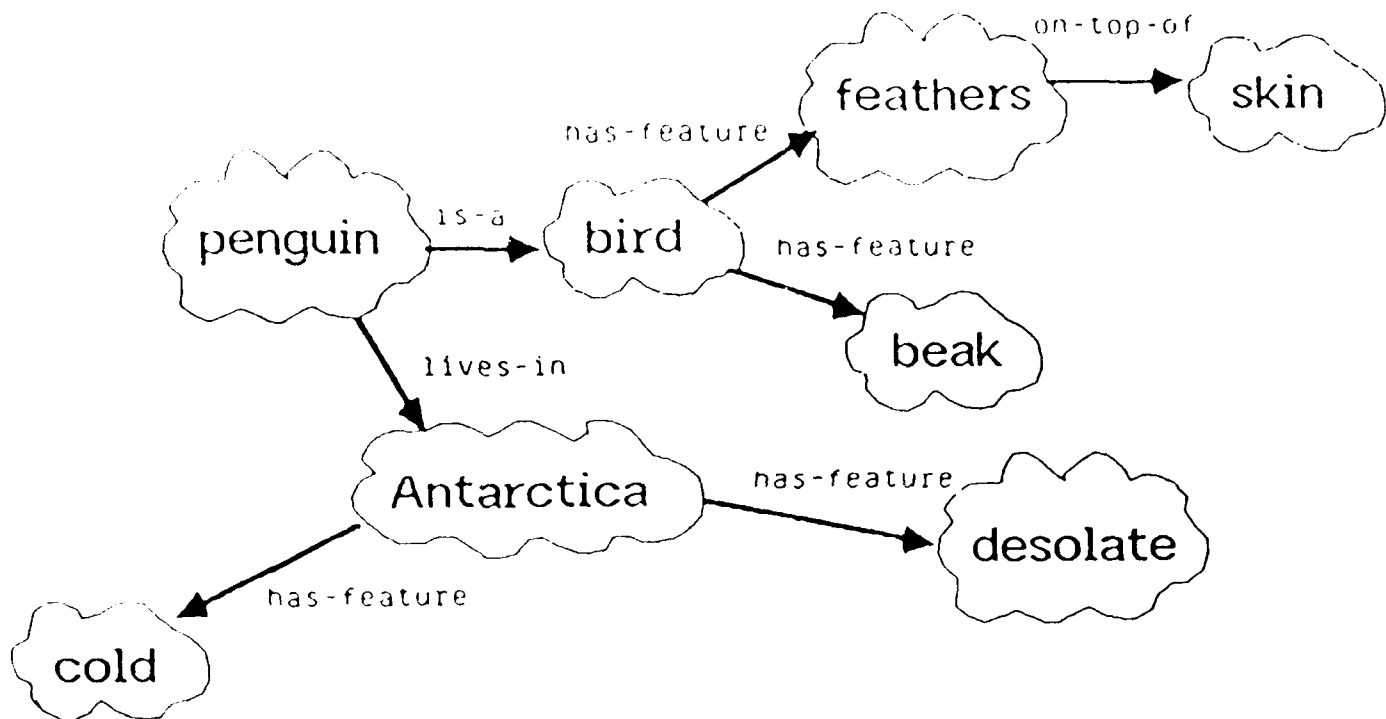
While linguists typically dismiss the idea of episodic memory as a "pragmatics" issue, artificial intelligence workers must grapple with the problems of pragmatics every day. The controversy in artificial intelligence between semantic and episodic memory goes far beyond natural language processing applications. Proponents of production-based expert systems are banking on semantic knowledge, while people who argue for case-based reasoning are committing themselves to episodic memory structures. Semantic knowledge provides a natural foundation for deductive reasoning (the programming language PROLOG is nothing more than a simple deductive mechanism operating over semantic knowledge bases). At the same time, some very interesting work is also being done on inductive reasoning within episodic memory [Lebowitz 1983b].

The question of underlying memory structures is important for any systematic study of nominal compounds. We have to account for the memory processes that tell us "chicken skin" makes sense. It is one thing to say that a semantic checking mechanism executes a test to see if chickens have skin. It is something else to specify the precise memory model that allows us to find correct answers. A complete process model for nominal compounds must therefore address complex issues of memory organization - it is not possible to finesse these problems (at least not knowingly) when the time comes to define an actual test function.

Memory Access and Knowledge Acquisition

A text in a specialized knowledge domain can be either easy to understand or difficult to understand depending on whether or not the reader has knowledge of the underlying domain. Terms like "base plate insulator welds" or "ship-shore tty sat communications" can render a text incomprehensible if the reader no conceptual foundation for the objects being described [Marsh 1983]. In many cases (especially for technical text), it is necessary to build new conceptual structures into memory at the same time a text is being processed. At first it may seem that the problem of memory access is then compounded by the problem of knowledge acquisition. But recent work in this area suggests that the knowledge acquisition problem goes hand-in-hand with memory access, and may even serve to simplify matters when it comes to complex noun phrases [Lebowitz 1983a, 1984a].

Does a penguin have skin?

penguin —is-a→ bird —has-feature→ feathers —on-top-of→ skin

bird —has-feature→ beak

penguin —lives-in→ Antarctica —has-feature→ desolate

Antarctica —has-feature→ cold

Does a chicken have skin?

dinner last night —main-course→ fried chicken ←as-food— chicken

dinner last night —location→ Kentucky Fried Chicken

Kentucky Fried Chicken —product→ fried chicken

fried chicken —favorite-part→ skin —can-be→ greasy

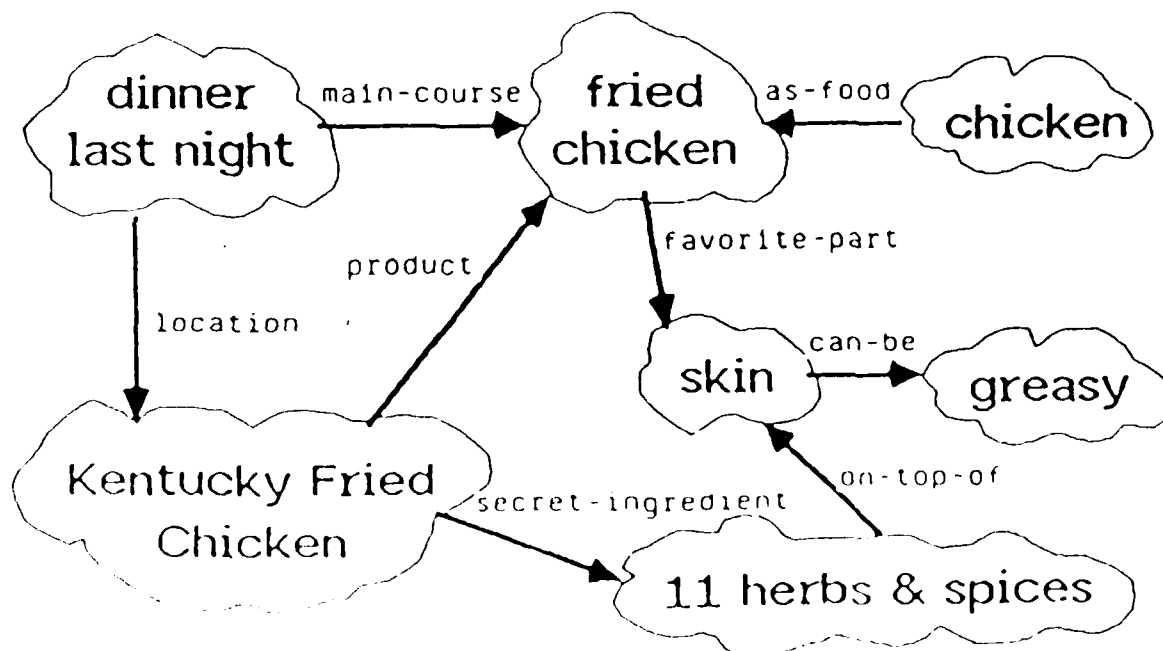Kentucky Fried Chicken —secret-ingredient→ 11 herbs & spices —on-top-of→ skin

Figure 1. Semantic Memory vs. Episodic Memory

Michael Lebowitz has implemented a system called RESEARCHER that maintains a knowledge base for patent abstracts describing computer hardware. This program has processed about 100 abstracts using a vocabulary of about 1200 words. Patent abstracts are characterized by a large number of long and complicated noun phrases, so RESEARCHER must maintain a rich knowledge base of relevant concepts and the relationships between them. What makes RESEARCHER unique is its ability to add knowledge to its memory as it reads new patent descriptions. This facility for knowledge acquisition makes RESEARCHER more robust in its text processing than than would otherwise be possible. The very processes of memory access can be changed over time to reflect newly-acquired concepts and generalizations created by RESEARCHER. The implications of this for the analysis of nominal compounds are dramatic. Lebowitz sums it up rather casually: " ... As with noun group processing, it is possible to develop heuristics that handle most cases, but they are complex and do not seem to be the right way to go for robust understanding."

Using techniques that Lebowitz first developed for newspaper stories [Lebowitz 1983a], RESEARCHER builds a "generalization-based memory" which is a form of inheritance hierarchy. The important point about this memory is that it automatically creates a hierarchy of object descriptions from examples. This type of memory provides a powerful foundation for semantic checking mechanisms during noun phrase analysis.

For example, suppose we are trying to understand the phrase:

magnetic read/write head enclosure

and we must decide whether "magnetic" is supposed to modify "read/write head" or "enclosure." Rather than test for semantic features or possible semantic features under each concept, we can examine the general concepts for a read/write head and an enclosure first, and then search through a tree of successively more specific concept instances until we find an example where one or the other is magnetic. This approach can only work if memory is sufficiently "fleshed-out," but the advantages seem worth the cost of some initial start-up overhead. In the event that no helpful instances can be located in memory, one could always fall back on heuristics of the sort described by Russell.

To understand better exactly how RESEARCHER uses memory to extract conceptual representations for nominal compounds, consider a simple noun-pair: "a motor spindle." Figure 2 shows a trace of RESEARCHER as it would process this phrase when there is no relevant information available in memory.

Running RESEARCHER at 3:19:26 PM

(A MOTOR SPINDLE)

Processing:

A   :New instance word – skip
MOTOR   :Memette within NP; save and skip
SPINDLE   :MP word – memette DRIVE-SHAFT#
New DRIVE-SHAFT# instance (&MEMO)
>>>Looking for relation between MOTOR# &MEMO (DRIVE-SHAFT#)
New MOTOR# instance (&MEM1)
Assuming &MEM1 (MOTOR#) and &MEMO (DRIVE-SHAFT#) are functionally related
Establishing UNKNOWN-PURP-REL relation; SUBJECT: &MEM1 (MOTOR#);
  OBJECT: &MEMO (DRIVE-SHAFT#) [&REL1]

Text Representation:

————————————A-0 MO DRIVE-SHAFT#

————————————A-1 M1 MOTOR#

A list of relations:

          Subject: Relation: Object:

[&REL1/A] &MEM1 (MOTOR#)   {UNKNOWN-PURP-REL} &MEMO (DRIVE-SHAFT#)

Figure 2. 'Motor spindle' with memory empty

Conceptual analysis begins after the head noun "spindle" is encountered. RESEARCHER then works backwards from the concept definition for a spindle (DRIVE-SHAFT#) to the concept definition for motor (MOTOR#) and tries to determine the relation (UNKNOWN-PURP-REL) between these two concepts. It searches memory for instances of drive shafts and motors, but finds no relevant examples. At this point RESEARCHER is forced to conclude that there is an indeterminate functional relation between these two objects. Now suppose we have relevant information available in memory. Specifically, suppose RESEARCHER had previously processed the noun phrase:

Ex 1: "a drive with a motor on top of a spindle"

Figure 3 shows the internal memory representation that RESEARCHER produces for this concept.

Text Representation:

```
        |———————A-1  MO  DRIVE#
————————0|————————A-2  M1  MOTOR#
———————————————————————M2  DRIVE-SHAFT#
```

A list of relations:

Subject:	Relation:	Object:
[&REL1/A] &MEM1 (MOTOR#)	{R-ON-TOP-OF}	&MEM2 (DRIVE-SHAFT#)

Figure 3. Setting up memory

Here we have an instance of a disk drive (DRIVE#) in memory with an "on top of" relation between the motor (MOTOR#) and the spindle (DRIVE-SHAFT#) used in the disk drive. If we now try to process "a motor spindle" with this information in memory, the conceptual analysis proceeds very differently. Figure 4 shows a trace of RESEARCHER's processing in this case.

Running RESEARCHER at 3:18:11 PM

(A MOTOR SPINDLE)

Processing:

A   : New instance word – skip
MOTOR   : Memette within NP; save and skip
SPINDLE   : MP word – memette DRIVE-SHAFT#
New DRIVE-SHAFT# instance (&MEM3)
>>>LOOKING for relation between MOTOR# and &MEM3 (DRIVE-SHAFT#)
New MOTOR# instance (&MEM4)
Establishing R-ON-TOP-OF relation; SUBJECT: &MEM4 (MOTOR#)
    OBJECT: &MEM3 (DRIVE-SHAFT#) [&REL2]

Text Representation:

————————————B-3   M3   DRIVE-SHAFT#

————————————B-4   M4   MOTOR#

A list of relations:

        Subject: Relation: Object:

[&REL2/B] &MEM4 (MOTOR#) {R-ON-TOP-OF}   &MEM3 (DRIVE-SHAFT#)

Figure 4.   'Motor spindle' with EX 1 in memory

22

When RESEARCHER looks for a relation between DRIVE-SHAFT# and MOTOR#, it finds the relevant example (figure 3) and assumes that the proper relation between a DRIVE-SHAFT# and a MOTOR# must be an "on top of" relation.

If RESEARCHER had previously processed a different example before encountering the phrase "a motor spindle," its conceptual analysis might have been different. For example, suppose that instead of hearing about "a drive with a motor on top of a spindle," RESEARCHER had previously encountered:

Ex 2: "a drive with a motor that includes a spindle"

In this case RESEARCHER would construct a part/assembly relation between the motor (MOTOR#) and the spindle (DRIVE-SHAFT#) as shown in Figure 5.

Running RESEARCHER at 6:40:31 PM

(A MOTOR SPINDLE)

Processing:

A   : New instance word – skip
MOTOR   : Memette within NP; save and skip
SPINDLE   : MP word – memette DRIVE-SHAFT#
New DRIVE-SHAFT# instance (&MEM3)
>>>Looking for relation between MOTOR# and &MEM3 (DRIVE-SHAFT#)
New MOTOR# instance (&MEM4)
Assuming &MEM3 (DRIVE-SHAFT#) IS PART OF &MEM4 (MOTOR#)

Text Representation:

```
——————————4|——————————3  M3 DRIVE-SHAFT#
                          M4 MOTOR#
```

A list of relations:

Subject:	Relation:	Object:
[none]	[none]	[none]

Figure 5. 'Motor spindle' with EX 2 in memory

As a result, RESEARCHER's interpretation of "a motor spindle" would now be based on this example, and the system would assume that a "motor spindle" must be a drive shaft contained in a motor.

As these last two examples illustrate, multiple interpretations for nominal compounds are possible in RESEARCHER. This raises the question of how the system should resolve conceptual ambiguities in the event that memory contains more than one relevant example, leading to more than one possible interpretation. Lebowitz believes that in such cases, the instances located closest to the "top of the tree" provide the best interpretations. Examples high in the tree structure correspond to the most general object descriptions, so this seems like a reasonable heuristic. For example, if a system has encountered 10 examples of oil trucks, one would expect that most of these instances would be describing trucks that carried oil. If one of the references described a truck that used oil for fuel, that example should be located further down in the tree structure since it is more specific (by virtue of the fact that there was only one). For more details on exactly how these trees are constructed, see [Lebowitz 1983b].

By accessing a dynamic memory structure that changes and grows the more it reads, we are much closer to simulating a system that learns domain expertise much as a person does. Some generalizations may be incorrect, and competing interpretations might be resolved differently depending on what the system has seen, but these imperfections can correct themselves as the system's exposure to its domain increases. Generalization-based memory may not be perfect, but does have the potential for being self-correcting [Lebowitz 1984b]. RESEARCHER is an exciting example of how episodic memory structures might be used to handle low-level problems in language analysis.

Conclusions

By examining a number of approaches to the problem of nominal compounds, we have seen a variety of memory models and representational devices proposed. Three distinct memory models have been described drawing from work in generative linguistics, modified formal semantics, and episodic memory modelling. By comparing and contrasting these three approaches, we have tried to emphasize the different research goals and theoretical premises that characterize each. The ostensible topic of interest has been nominal compounds throughout. But individual treatments of this topic vary considerably across theoretical frameworks. From an interdisciplinary vantage point, there seems to be a clearly defined sense of evolution and direction. We are moving from a purely linguistic paradigm that manages to avoid the problem internal meaning representation altogether, toward a process model paradigm that embraces not only the problems of conceptual representation but global memory organization as well. This trend may not be welcomed or even acknowledged by all

24

concerned, but it is an inevitable fact for those of us who wish to design detailed process models for natural langauge processing. If we are willing to confront language in terms of cognitive facilities, language processes become inseparable from memory processes. This perspective on language can entice us with great promise but it probably boasts the greatest difficulties as well. Let us hope that it inspires our most diligent research efforts in return.

Bibliography

Becker, J. D. (1975) "The phrasal lexicon" in *Theoretical Issues in Natural Language Processing*. M.I.T. Cambrige, MA. pp. 60-63

Bransford, J.D. & Franks, J.J. (1971) "The abstraction of linguistic ideas" *Cognitive Psychology* 1971, 2, pp. 331-350.

Charniak, E. (1981) "A Parser with Something for Everyone" Technical Report No. CS-70, Department of Computer Science, Brown University. Providence, RI.

Coker, P. & Crain, S. (1979) "Lexical access during sentence processing" a paper presented at the Linguistic Society of America Annual Meeting. University of California at Irvine.

Crain, S. & Coker, P. (1979) "A semantic constraint on parsing" a paper presented at the Linguistic Society of America Annual Meeting. University of California at Irvine.

DeJong, G. (1979) "Prediction and substantiation: a new approach to natural language processing" *Cognitive Science* 3, 1979, pp.251-273.

Dyer, M. (1983) *In-Depth Understanding: A Computer Model of Integrated Processing for Narrative Comprehension*. MIT Press. Cambridge, MA.

Finin, T. W. (1982) "The interpretation of nominal compounds in discourse" Technical Report MS-CIS-82-3, Moore School of Engineering, University of Pennsylvania.

Gershman, A. V. (1979) "Knowledge-Based Parsing" (PhD thesis) Research Report #156, Department of Computer Science, Yale University. New Haven, CT.

Jespersen, O. (1942) *A Modern English Grammar on Historical Principles*, Copenhagen: Ejnar Munksgaard.

Katz, J.J. & Fodor, J. A. (1964) "The structure of a semantic theory" in *The Structure of Language* (ed. Foder/Katz), Prentice-Hall, Inc. Englewood Cliffes, NJ.

Koziol, H. (1937) *Handbuch der Englischen Wortbildungslehre*. Heidelberg: Carl Winter's Universitatsbuchhandlung.

Lebowitz, M. (1983a) "Memory-Based Parsing" *Artificial Intelligence* 21, 4, pp. 363-404.

Lebowitz, M. (1983b) "Generalization from natural language text" *Cognitive Science* 7,1, pp. 1-40.

Lebowitz, M. (1984a) "Using Memory in Text Understanding". Technical Report, Department of Computer Science, Columbia University.

Lebowitz, M. (1984b) "Conceptual processing when things go wrong" *Cognition and Brain Theory* vol. 7, no. 3-4, pp. 375-398.

Lees, R.B. (1960) "The grammar of English nominalizations" *International Journal of American Linguistics* 26, Publication 12.

Lehnert, W. (1978) *The Process of Question Answering*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Lehnert, W. (1979) "Representing Physical Objects in Memory" in *Philosophical Perspectives on Artificial Intelligence* (ed: M. Ringle) Humanities Press, Atlantic Highlands, NJ.

Lehnert, W. (1982) "Plot units: a narrative summarization strategy" in *Srategies for Natural Language Processing* (eds: Lehnert/Ringle). Lawrence Erlbaum Associates, Hillsdale, NJ. pp. 375-414.

Lehnert, W., Dyer, M., Johnson, P., Yang, C.J., Harley, S. (1983) "BORIS - An experiment in in-depth understanding of narratives" *Artificial Intelligence*, vol. 20 1983. pp. 15-62.

Lehnert, W. (1984) "Narrative Complexity based on summarization algorithms" in *Computational Models of Natural Language Processing*. (eds: Bara/Guida) Amsterdam: North-Holland. pp. 247-259.

Lehnert, W. and Rosenberg, S. (1985) "The PLUM Users Manual" CPTM #1, Department of Computer and Information Science, University of Massachusetts. Amherst, MA.

Levi, J. (1978) *The Syntax and Semantics of Complex Nominals*. Academic Press, New York.

26

Lytinen, S. (1984) "The organization of knowledge in a multi-lingual, integrated parser." (PhD thesis) Research Report #340. Department of Computer Science, Yale University. New Haven, CT.

Marchand, H. (1969) *The Categories and Types of Present-Day English Word Formation*, (2nd edition) Munich: C. H. Beck'sche Verlagsbuchhandlung.

Marcus, M. P. (1980) *A Theory of Syntactic Recognition for Natural Language*. MIT Press, Cambrigde, MA.

Marsh, E. (1983) "Utilizing domain-specific information for processing compact text" in the *Proceedings of the Conference on Applied Natural Language Processing*, Irvine, CA. pp. 99-103.

Milne, R. W. (1982) "Predicting garden path sentences" *Cognitive Science*, vol. 6, no. 4, pp. 349-373.

Riesbeck, C. K. & Martin, C. E. (1985) "Direct memory access parsing" Research Report #354, Department of Computer Science, Yale University. New Haven, CT.

Riesbeck, C. and Schank, R. (1976) "Expectation-Based analysis of sentences in context" Research Report #78, Department of Computer Science, Yale University. New Haven, CT.

Russell, S. W. (1972) "Semantic categories of nominals for conceptual dependency analysis of natural language" Stanford Artificial Intelligence Project MEMO AIM-172, STAN-CS-72-299. Computer Science Department, Stanford University.

Schank, R. (1975) *Conceptual Information Processing*. American Elsevier, New York.

Schank, R. and Abelson, R. (1977) *Scripts, Plans, Goals and Understanding*, Lawrence Erlbaum Asociates, Hillsdale, NJ.

Small, S. (1980) "Word expert parsing: A theory of distributed word-based natural language understanding" (PhD thesis) TR-954, Department of Computer Science, University of Maryland.

Tulving, E. (1972) "Episodic and semantic memory" in *Organization of Memory* (eds: Tulving/Donaldson), Academic Press, New York.

Waltz, D. L. and Pollack, J. B. (1984) "Phenomenologically plausible parsing" in *Proceedings of the 1984 American Association for Aritificial Intelligence Conference*, pp. 335-339.

Wilensky, R. (1981) "A knowledge-based approach to language processing" in *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, pp. 25-30.

Wilks, Y. (1972) *Grammer, Meaning, and the Machine Analysis of Language*, Routledge & Kegan Paul, London.

END

1-87

DTIC