INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the original text directly from the copy submitted. Thus, some dissertation copies are in typewriter face, while others may be from a computer printer.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyrighted material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is available as one exposure on a standard 35 mm slide or as a 17" × 23" black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. 35 mm slides or 6" × 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
English grammar as a stratified system of signs

Edmiston, Cynthia Denise, Ph.D.
Rice University, 1988
RICE UNIVERSITY

ENGLISH GRAMMAR AS A
STRATIFIED SYSTEM OF SIGNS

by

CYNTHIA D. EDMISTON

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

DOCTOR OF PHILOSOPHY

APPROVED, THESIS COMMITTEE:

Sydney M. Lamb, Professor of Linguistics and Semiotics, Director

James E. Copeland, Professor of Linguistics and Semiotics

Philip W. Davis, Professor of Linguistics and Semiotics

Robert Cartwright, Professor of Computer Science

Houston, Texas

May, 1988
ABSTRACT

ENGLISH GRAMMAR AS A STRATIFIED SYSTEM OF SIGNS
Cynthia D. Edmiston

Linguistic information, representing the knowledge that a speaker has of his or her language, can be uniformly represented as a system of signs. In addition to the traditional linguistic signs, the connections between morphemes or morphemic words and their meanings, syntactic constructions, idiomatic expressions, discourse phenomena, all types of linguistic information, are treated as meaningful and therefore capable of being represented as signs.

A four-strata version of linguistic structure is adopted, with graphemic, morphemic, lexicemic, and sememic levels. The signs mediate between the information at the different levels, such that the expression and content of each sign is "local", and expression and content in the more general sense (all the way from graphemes to sememes) are related indirectly.

The signs which constitute a core grammar of English are formalized and set forth. This formalization, the major goal of the dissertation, is not simply a description of texts, however. Rather, it is an attempt to construct a cognitive model which accounts for those texts.

Two points of theory are: (1) there is a distinction between relations (signs) and processes (encoding and decoding), and (2) language is an adaptive system subject to continuous modification. The linguistic information in the
A semiotic format is shown to be useable to produce and decipher texts, and to be readily modified when new information is encountered.
ACKNOWLEDGEMENTS

I would like to acknowledge the role of my family in the completion of my doctorate, and thank them for their constant support and encouragement: my parents, James Aubrey and Pauline Syphrett Edmiston, for their confidence in me and their financial aid throughout these many years of school; my sister, Gretchen Edmiston, for being the most generous and encouraging person I know; my brother-in-law, Robert Merrill, for his kindnesses; and especially my sweetheart, David Janowitz, for his caring and generosity.

I am most grateful to The Mitchell Trust for providing financial support in the form of a fellowship for the years 1983-8, and to the Linguistics Department of Rice University for summer research grants in 1984 and 1986.

I would like to extend my sincere thanks to Professor Sydney Lamb, my dissertation advisor, for all of his patience, help, and guidance, not only during the writing of this dissertation, but also throughout my five years at Rice. I am also grateful to Professor James Copeland, a truly equitable and encouraging person. I thank Professor Philip Davis for his teachings and advice, Dr. Lilly Chen for being the "friendly spirit" at all of my oral examinations, and our department secretary, Sandra Gilbert, for her always cheerfully given help. I would also like to thank Professor Robert Cartwright for serving as a member of my committee. I am very appreciative to Professor Douglas
Mitchell for his wonderful sense of humor and his inordinate patience in preparing me for my German reading exam.

I wish to thank each of the linguistics graduate students for their camaraderie, making my years at Rice a time I will always remember fondly. And to my special friends, Chang-in Lee, Lillian Huang, Pete Cobin, and Katharina Barbe, a heartfelt thanks for all the fun, foolishness, help, and friendship.

I would also like to thank my two good friends whose companionship has kept me going, Kate Pilgrim and Susan Lamb, and my ballet teachers, Ms. Soili Arvola and Mr. Leo Ahonen, who worked on my body while Rice worked on my mind.

Finally, I wish to thank Max and Plum Pudding for supplying me not only with numerous examples to illustrate the points in my dissertation, but also countless hours of entertainment.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>General remarks</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>The linguistic sign</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>Terminology</td>
<td>17</td>
</tr>
<tr>
<td>1.3</td>
<td>Previous work</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Notes</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>SYSTEM OVERVIEW</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>General remarks</td>
<td>24</td>
</tr>
<tr>
<td>2.1</td>
<td>Basic system design</td>
<td>25</td>
</tr>
<tr>
<td>2.2</td>
<td>Notation for representing linguistic information</td>
<td>29</td>
</tr>
<tr>
<td>2.3</td>
<td>The dynamic aspect</td>
<td>41</td>
</tr>
<tr>
<td>2.4</td>
<td>The procedure for encoding and decoding</td>
<td>43</td>
</tr>
<tr>
<td>2.5</td>
<td>System applications</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Notes</td>
<td>52</td>
</tr>
<tr>
<td>3</td>
<td>THE MORPHICON</td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>General remarks</td>
<td>54</td>
</tr>
<tr>
<td>3.1</td>
<td>Nouns</td>
<td>57</td>
</tr>
<tr>
<td>3.2</td>
<td>Determiners</td>
<td>60</td>
</tr>
<tr>
<td>3.3</td>
<td>Pronouns</td>
<td>63</td>
</tr>
<tr>
<td>3.4</td>
<td>Numerals</td>
<td>65</td>
</tr>
<tr>
<td>3.5</td>
<td>Adjectives</td>
<td>66</td>
</tr>
<tr>
<td>3.6</td>
<td>Adverbs</td>
<td>68</td>
</tr>
<tr>
<td>3.7</td>
<td>Prepositions</td>
<td>70</td>
</tr>
<tr>
<td>3.8</td>
<td>Verbs</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Notes</td>
<td>76</td>
</tr>
<tr>
<td>4</td>
<td>THE LEXICON</td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>Entries in the Lexicon</td>
<td>77</td>
</tr>
<tr>
<td>4.1</td>
<td>Idioms</td>
<td>77</td>
</tr>
<tr>
<td>4.2</td>
<td>Number (Plurality)</td>
<td>79</td>
</tr>
<tr>
<td>4.3</td>
<td>Adjective phrases</td>
<td>79</td>
</tr>
<tr>
<td>4.4</td>
<td>Nominals</td>
<td>81</td>
</tr>
<tr>
<td>4.5</td>
<td>Noun phrases with specific reference</td>
<td>83</td>
</tr>
<tr>
<td>4.6</td>
<td>Noun phrases with generic reference</td>
<td>87</td>
</tr>
<tr>
<td>4.7</td>
<td>Other noun phrases</td>
<td>89</td>
</tr>
<tr>
<td>4.8</td>
<td>Noun attributes</td>
<td>90</td>
</tr>
<tr>
<td>4.9</td>
<td>Prepositional phrases</td>
<td>92</td>
</tr>
<tr>
<td>4.10</td>
<td>Tense</td>
<td>92</td>
</tr>
<tr>
<td>4.11</td>
<td>Marked word order of questions</td>
<td>94</td>
</tr>
<tr>
<td>4.12</td>
<td>Some &quot;hopping&quot; (anataxis)</td>
<td>97</td>
</tr>
<tr>
<td>4.13</td>
<td>Subject verb agreement</td>
<td>98</td>
</tr>
<tr>
<td>4.14</td>
<td>Aspect and voice</td>
<td>100</td>
</tr>
<tr>
<td>4.15</td>
<td>Negative</td>
<td>100</td>
</tr>
<tr>
<td>4.16</td>
<td>Verbs</td>
<td>101</td>
</tr>
</tbody>
</table>
Chapter 1
INTRODUCTION

1.0 General Remarks

The purpose of this dissertation is to test the hypothesis that linguistic information representing the knowledge that a speaker has of his or her language can be uniformly represented as a stratified system of signs. The data for the investigation is a core grammar of English whose formal specification is based on the new "dynamic grammar" system proposed by Lamb (in preparation) for language description. One of the requirements for a dynamic grammar is that the linguistic information be organized in a dictionary-like format, which can be used either by human beings or computers. Each entry in the grammar dictionary corresponds to a linguistic sign as defined in dynamic grammar and Cognitive Linguistics.

A linguistic model, to be cognitively realistic, must be usable to produce and understand texts. To show that the linguistic information in these dictionary entries can be used in the processes of production and understanding, some traces of the encoding and decoding of example sentences are presented. In addition, since humans update their linguistic information in the process of using it (for example, unfamiliar lexical items are added when they are encountered), a few techniques for the modification of information are outlined informally.
As an introduction to both the formal description of the linguistic information and the informal sketch of its use in processing, background information about the concepts on which this new description rests is provided in 1.1 while a brief explanation of a few technical terms is presented in 1.2. Some remarks about previous work appear in 1.3. The intention is to furnish the uninitiated reader with sufficient information to elucidate the approach to the problem at hand so that he will not have to seek out references to gain a basic understanding.

A detailed description of how the whole system is organized is then undertaken in Chapter 2. This overview of the system includes discussion on its basic design, the entries that contain the linguistic information and how to interpret them, aspects of system modification, decoding and encoding procedures, and some applications for such a system.

The main task of this dissertation is to lay out the linguistic information in dictionary-like entries, the descriptive framework that Lamb has proposed. This is done in Chapters 3, 4, and 5. A complete grammar of English, however, is not attempted. First, such a grammar is far beyond the scope of one dissertation, and second, a well-circumscribed, fixed grammar is an impossibility within the theoretical framework being advanced in this treatise wherein language is viewed as a system that is constantly being updated. Such basic parts of the grammar as simple
declaratives, wh-questions, yes-no questions, and complex sentences with phrasal and clausal attribution are being covered. Entailed by these sentence types are the so-called inverted word order of questions; subject verb agreement; pronoun antecedent agreement; the verb phrase with adverbs, negatives, tense, modality, aspect, and voice; the noun phrase with determiners, numerals, and adjectives and other attributives such as relative clauses, prepositional phrases, and participial phrases; event valency; criteria for types of participants; circumstantial of time, location, manner, and so forth; and essential discourse information such as focus.

Chapters 6 and 7 are devoted to showing how the linguistic information can be used and updated in production and understanding.

The processes of production and understanding, together with the linguistic information and the ability to modify that information, constitute a whole system. That which is presented in this dissertation is a partially formalized, abstract model of such a system, which is intended to be implemented, when fully formalized, as an automated system for natural language processing.

1.1 The linguistic sign

Sydney Lamb's work of the 1960's was referred to as Stratificational Linguistics, and even now that title may be the one most generally identified with Lamb's name although he began referring to his investigation as fitting into the
broader scope of cognitive linguistics in the early 1970's (Lamb 1971). To be sure, many of the basic tenets of stratification theory may be found in Lamb's current work, but the broader term 'cognitive' better encompasses the range of linguistic questions which Lamb's evolved philosophy embraces.

Cognitive Linguistics attempts to model the information system that is present in the mind of the typical (not the ideal) individual. It differs sharply both from Descriptive Linguistics, which attempts to describe the structures of texts and collections of texts, primarily through the use of processes of segmentation and classification, as well as from Generative Linguistics, which attempts to characterize, by means of formal rules, the set of possible sentences or texts that the ideal speaker is able to produce. (Lamb 1986: 55)

The individual's information system, the focus of Cognitive Linguistics, consists of a network of relationships, the basic organizing unit of which is the 'nection' (cf.'con-nection') (Lamb 1966:49-52 and 1984:76-7).

Recently, Lamb (1984,1986) has compared the nection to the 'sign', discussed early in this century by Ferdinand de Saussure, and found the nection to be a sign-like unit. He has referred to the nection as an internal sign or a micro-sign because it designates that which is in the individual's, as opposed to the community's, system. (De Saussure had characterized language as the system of signs shared by a speech community.) Like de Saussure's sign, the nection is the association between a signified, or content, and a signifier, or the expression for that content. The ways in
which the sign-like nection differs from the traditional concept of the sign are significant, however, and are explored in that which follows.

De Saussure represents his version of signs as '... real objects; ... they are the concrete entities of our science.' (de Saussure 1916. in 1966:102) The sign is defined by its separation from everything that surrounds it. 'These delimited entities or units stand in opposition to each other in the mechanism of language.' (de Saussure 1916. in 1966:103) De Saussure offers visual images of his concept of a sign as shown below.

(signified

signifier)

(concept

sound-image)

(tree

arbor)

(de Saussure 1916. in 1966:114,66,67)

Although de Saussure warns his readers to consider the sign as a form and not a substance (de Saussure 1916. in 1966:113,122), many subsequent linguists have viewed it as a substantive object. The Structuralists or Descriptivists of the 1920's - 1950's were somewhat concerned with this 'item', but '... the notion of sign was generally taken as applying only to the morphemes of a language. Everything else -- phonology, morphology, syntax, and whatever else may have been recognized by a given theory -- was evidently part of the system by which the symbols (the morphemes) were combined or otherwise operated upon; and most of the attention was given to those other aspects of the structure,
however conceived ...' (Lamb 1986:53).

Generative linguists, of course, do not endorse the notion of language as a system of signs. Rather, language is the set of sentences that may be generated by the syntactic rules, so syntax has been their focus (cf. 2.1).

Cognitive linguists do subscribe to the concept of the sign, but to a more refined, flexible version than that of their predecessors. The notion is '... a relation, not an object. It is a relation between a signified and a signifier.' (Lamb 1986:56) A relation may be illustrated by a connecting line between the signified and the signifier as shown below.

```
  signified   "tree"
     |           |
signifier    tree
```

'We cannot think of the internal sign in the way that would come from a literal interpretation of Saussure's famous diagram which depicts the sign as an oval object made up of two sides. That concept of the notion would break down as soon as we consider synonymy; that is, a single signified but different signifiers. The view of the sign as an object requires two separate such objects, with the signified duplicated.' (Lamb 1986:56-7) The following diagram illustrates such duplication.
We would get even more duplication, far more if we consider the question of what the signified really is ... Let us think about the meaning of 'cat'. It includes what a typical cat looks like, a visual image; it includes what a meow sounds like, an auditory image; it includes the fact that cats have fur, that cats have tails, that a cat is a common type of domestic animal, a pet, that is has four legs; all of these and more constitute part of the meaning of 'cat'. But most of these properties are also aspects of the meaning of many other items. The meaning of 'fur' is not only an aspect of the meaning of 'cat', it's also an aspect of the meaning of 'dog', 'bear' and so on. Meanings do not in general consist of just one or two isolated components. Now if we try to model the meaning side of the sign as an object, then the only way to get a realistic description is by having an enormous amount of duplication. Components like ... 'fur' will have to be duplicated hundreds, perhaps thousands of times.

The alternative is to represent the notion not as a chunk of meaning [signified] and a chunk of expression [signifier], but rather, as a relationship. Then we only have the concept ... ['fur'] once in the conceptual system, and it will have connections, direct or indirect, to everything that has ... [having fur] as a part of its meaning... And so the notion has to be seen as a relational entity, as a connection between a unit of content and a unit of expression. And the system of notions ... is a network of relationships. (Lamb 1986:57-8)

One might wonder what harm there is in having excessive duplication in the system. After all, the human brain certainly has the capacity to accommodate inordinate amounts of information. The repetition implies, however, that each duplicated instance of the concept is separate from the
other instances, and that is a misrepresentation. The concept occurs only once since it is the same concept for all of the expressions that represent it.

Another way the nection differs in theory from the traditional sign is in what it serves to represent. For de Saussure, all of the signs together comprise the language of the speech community. For Lamb, on the other hand, all the nections together constitute the linguistic system of the individual. How does this distinction affect the make-up of the sign or nection?

For de Saussure, language exists only in the collectivity, as a public domain, and is not an individual matter. The signs are, therefore, fixed entities; the signifieds and signifiers and their associations being identical for each member of the speech community. Signs appear to be foregone conclusions, just waiting for the individual to come along and learn and use them. 'No individual, even if he willed it, could modify in any way at all the choice that has been made [for a sign] ...; it is bound to the existing language.' 'No matter what period we choose or how far back we go, language always appears as a heritage of the preceding period.' (de Saussure 1916.in 1966:71) 'Language is checked not only by the weight of the collectivity, but also by time. These two are inseparable. At every moment solidarity with the past checks freedom of choice.' (de Saussure 1916.in 1966:74)

To a certain extent, de Saussure's claims are true.
The individual, in his social setting, learns, more or less, the conventional signifiers which have been established for certain concepts. To operate by making up his own signs would render him dysfunctional in society. To assert, however, that language is somehow a perfected system, shared identically by all and unalterable by the individual, places language outside of the human being. However, language does not exist outside of the humans who use it, and humans by their very natures do not possess neat, fixed systems which are exactly like everyone else's.

Language is not a perfect communication system such that everyone shares exactly the same concepts and expressions for those concepts. A person's internal network of relationships, his collection of signs as it were, is the result of his personal experiences. Whatever connections there are in an individual's system reflect that person's apprehension of (interaction with) the world around him. Because people are human and live together on the planet and interact socially with one another, they have similar signifieds, signifiers, and interconnections, but to assume that all of their signs completely overlap is to ignore what it means to be human, that individual experience colors perception and ideas. For example, most people have a signifier 'tractor' in their systems. One might assume that, across individuals, the signified for 'tractor' might be 'machine used to work dirt'. (Of course, it is possible that some people do not have even that much information about a
tractor.) A truck farmer, however, has a great deal of additional information which takes part in the make-up of his concept of tractor that an urban dweller, who has perhaps only seen a tractor in his first-grade reader or upon passing a highway-building project, does not. As Lamb points out, '... to the extent that different individuals are able to communicate, it is by virtue of ... correspondences, however imperfect, among modules of their individual systems.' (1986:56)

To study the typical individual's system of relationships is then much simpler (Lamb 1986:54) than trying to project for an entire community some common denominators for associating signifiers with signifieds. It may well be the case that no such common denominator exists for an entire community. As James Sledd has pointed out in his study of dialects, '...people who rarely talk together ... talk differently ...' (Sledd 1969:1307). Some might be willing to dismiss such differences as unimportant, but it is erroneous and naive to assume that people understand one another perfectly, that they share one system, even when ostensibly they speak the same language. Understanding is imperfect at best, and often wholly absent. This lack of understanding does not issue from an ignorance of the meaning of particular words, but rather from a signifier's having varying signifieds in different individuals.

Recognizing that there are differences between (1) the individual's system, (2) a system shared by a community
(points of intersection among individuals' systems), and (3) a system that is the composite of all individuals' systems might cause one to wonder just which of these phenomena should be classified as a language.

Examining first the individual's system, can one give a language name like English or French to what is found? The answer must be no because a bi- or multi-lingual individual has in his system that which one might refer to as many different languages. So, a language cannot be the individual's system. For the same reasons, a language designation cannot be assigned to the composite of all individuals' systems, the third choice mentioned above.

That leaves the second choice, the intersection of individuals' systems, something similar to that which de Saussure called a language. As has been shown, however, all individuals may not share common linguistic signs. While it would be abnormal for an individual to share no points of intersection with his neighbors, it is extremely common, in fact the rule rather than the exception, that he not share some, or even many such points. How then can one assign a language label to some set of common denominators when they are not indeed common at all?

This leaves one in the uncomfortable position of having to admit that what has always been reified as an existing entity, a language, may in reality be only an illusion, a label assigned by the individual to help him conceptualize and communicate.
Lamb offers an alternative. He suggests that, instead of trying to define 'a language', a count noun, one define 'Language', a mass noun. Language is simply the composite of networks of relationships that all individuals possess (choice three, above). Choice one, the individual's linguistic system, can also be called Language, or perhaps more appropriately, "some" Language. Choice two, above, the intersection of individuals' systems fails because it does not include all of the notions or relationships that a given individual has.

Another difference between the traditional sign and the notion lies in what they connect. De Saussure's sign joins that which is primarily a word or morpheme to the concept associated with it. The notion '...in addition to modeling the ordinary, usual linguistic signs, the ones that are like morphemes, ... [is] also ... at other levels of structure, both above and below the morpheme.' (Lamb 1986:58)

For de Saussure, the signified is always a concept. This restriction is enough to handle all of language, however, since de Saussure's linguistic structure has only one level, the grammatical, which embraces syntax, morphology, and lexicology (de Saussure 1916.in 1966:135). Syntax and the '-ologies' are taken care of by devices other than the sign, and phonology is considered by de Saussure to be external to language (de Saussure 1916.in 1966:33).

The cognitive view of linguistic structure as stratified admits several levels, including phonology, and by its
very nature demands a broader interpretation of the sign. The signified in this theory is not restricted to concepts. A signifier may have 'local meaning', and it may be something other than a word or morpheme.

Looking ... at lower levels [of linguistic structure], we have first what we might call the low-level morpheme. For example, let us consider 'took', as in 'they took the books'. What is 'took'? It looks a little like a morpheme, in fact just like a morpheme on the expression side. But its upward connections are not to some meaning. They are just to the morpheme 'take' and the morpheme 'past tense', which are already there in the system. We don't have to duplicate the meaning of 'take' in 'took'. All we have to do is connect 'took' upwards to 'take', which is there anyway. This is a nection whose content is not meaning in the usual sense. Its content is just connection to a couple of morphemes. Its meaning is what it has by virtue of those two connections. [This is 'local meaning'.]

And then, looking below this level, we get into the phonological area, where we have nections whose content is just connection to various morphemes. What is the traditional phoneme, when interpreted cognitively? There are two ways of looking at the phoneme. One, it is the unit which distinguishes morphemes one from another. That's the upward side of the phoneme. Its meaning, if you will, is its participation as a component in various morphemes. This is what meaning is at the phonological level. And the lower side is the phonetic expression. So a phoneme is also a sign-like unit, like every other part of a linguistic structure. (Lamb 1986:59-60)

...Looking above the morpheme, we have nections like 'blackboard'. On the expression side, we don't have to connect 'blackboard' all the way to units of the phonological level; it just connects to 'black' and 'board', both of which are morphemes, present anyway in the system; this kind of treatment is easy to handle if we use a relational approach. This type of signifier is not in itself phonological, but morphological; the expression side of 'blackboard' is connected not to phonological units but to 'black' and 'board', which are morphemes. Thus 'blackboard' is a higher level nection, one whose expression side is
morphemes rather than phonological units. (Lamb 1986:58-9)

De Saussure's syntax is a syntagmatic relation between signs. 'Combinations supported by linearity are syntagms.' (de Saussure 1916 in 1966:123) The nature of syntax for Lamb, however, is sign-like, just as it is for elements like morphemes, phonemes, lexemes, etc.

We also find syntactic nections. The usual approach to syntax involves analyzing sequences of words such as sentences and devising rules to generate them. I would like to suggest that this approach misses the point. The point about syntax, without which it can never be properly treated, is that syntactic constructions are meaningful; that's what makes them what they are. That is the property that provides the explanation for the existence of constituent structure and of syntactic phenomena in general. This healthy approach is what we find not only in Cognitive Linguistics but also in Functional Linguistics. Instead of seeing syntax as a list of rules ..., we should see syntax as a collection of individual syntactic constructions, each of which is concerned, on the one hand, with a particular syntactic property or specific combination, and on the other hand, with a meaning. Whatever the syntactic device is, that is the expression that is used for expressing some meaning. ...Thus each syntactic construction in the individual linguistic system is a nection... (Lamb 1986:59)

This view of the internal linguistic structure of the individual, then, involves interconnected nections at various levels, all of them relational; so that the system as a whole is a network of relationships ..., but a network made up of numerous small modules, each of which is a semiotic unit... This view contrasts sharply with the notion of language as an object of some kind. (Lamb 1986:60)

That nections are not the same, or fixed, for all members of a speech community has already been mentioned. The nection is variable in another respect as well. Whatever one chooses to relate or link is dependent on his
experience, and since normal humans continue to grow and learn throughout their lives, they continue to add and modify the relations they have previously made. People's linguistic systems are not static, then. The network expands and is refined constantly as people interact with their environment. Conceptual structure is especially dynamic in this respect.

The language of the individual, the individual linguistic system, is nothing but an interconnected system of nections, and it continually expands as new nections are added. There are certain periods in which nections of certain levels are added with great frequency. But there is no period of life in which they stop being added, except perhaps in some cases of senility. (Lamb 1986:60-1)

Given the cognitive view of language described thus far, one might wonder how language fits into the broader realm of general cognitive systems. One must realize that the nection is not peculiar to language, but is '... the basis of cognitive semiotic structures in general.' (Lamb 1986:61) Indeed, linguistic nections clearly have extralinguistic connections; for example, the visual and auditory connections that enter into a person's concept of 'cat' as mentioned at the beginning of this section. 'As the linguistic system is so inextricably intertwined with other portions of the general semiotic system, interconnecting and overlapping with all of its other modules, it appears that the notion sometimes proposed of a separate language faculty, as distinct from general human intelligence, is an illusion.' (Lamb 1986:62)
Summary

The focus of investigation for Cognitive Linguistics is the linguistic system of the typical individual. That system is the network of relationships made up of nections that the individual has amassed through personal experiences, both linguistic and other, and it differs from the next individual's network precisely because of those experiences.

This notion of the sign-like unit, the nection, differs from de Saussure's sign in that the association between the parts of the unit, the signified and the signifier, is not one of coupling and isolating the parts so that a delimited entity exists, but of relating the parts and allowing those same parts to relate to other parts so that an interlocking network results. A nection is not fixed across individuals; that is, an individual may have connections between signifiers and signifieds different from his neighbor. Neither is a nection fixed in the sense that it can never be changed once it is formed. Humans continually learn and refine their linguistic systems. A nection is not simply a connection between some semantic meaning and a word or morpheme. It is the link anywhere in the system between an element of local expression and its local meaning. Any element of local expression (phoneme, morpheme, word, idiom, syntactic construction, etc.) has connections upward (toward, but not necessarily to, except indirectly, semantic meaning), and those connections are to the local content (signified) for that expression. The expression, in turn, may serve as the
local content for some other expression further down (to-
ward, but not necessarily to, the sounds) in the structural
network.

Finally, "... there is evidently no such thing as a
'language faculty' as distinct from general human intelli-
gence, which can properly be understood as a general semio-
tic faculty." (Lamb 1986:62)

1.2 Terminology

Each of the four levels (strata) of linguistic
structure has an inventory of elements and a set of
combinatory possibilities (co-occurrence restrictions) for
those elements called a tactics. The following table shows
the names of the levels, elements, and tactics, as well as
the major unit of structure that each tactics specifies.
(Cf. Lamb 1966 and Lockwood 1972 for complete, formal
definitions of the following terms.)

<table>
<thead>
<tr>
<th>Level</th>
<th>Elements</th>
<th>Tactics</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>sememic</td>
<td>sememes</td>
<td>semotactics</td>
<td>clause</td>
</tr>
<tr>
<td>lexemic</td>
<td>lexemes</td>
<td>lexotactics</td>
<td>phrase</td>
</tr>
<tr>
<td>morphemic</td>
<td>morphemes</td>
<td>morphotactics</td>
<td>word</td>
</tr>
<tr>
<td>phonemic</td>
<td>phonemes</td>
<td>phonotactics</td>
<td>syllable</td>
</tr>
</tbody>
</table>

There are two mediums of expression for language,
speech and writing. The phonemic level deals with spoken
language while the graphemic level, its alternant, deals
with written language. It is this alternant which is
considered in the present treatment of the grammar (cf.2.1).

By convention, the sememic (conceptual) level is
characterized as the highest or deepest level, and the phonemic/graphemic level is the lowest or most surface level. When comparing the morphemic and graphemic levels of analyses, for example, a statement such as 'the higher of the two levels of analysis' is understood to mean the morphemic level.

The relationship between elements on the same level is tactical. A tactics (for example, the morphotactics) states the co-occurrence restrictions for the elements (the morphemes) such that well-formed units (words) are specified. The semotactics specify well-formed clauses, another term for which is "proposition".

The relationship between elements on different levels is realizational. For example, a lexeme is "realized as" or is the "realize of" a morpheme; or, a morpheme "realizes" or is the "realization of" a lexeme.

Some particular types of realization referred to in the body of this work are "empty realization" in which a lower level element has no overt realize, "zero realization" in which an upper level element is realized by no lower level element, "portmanteau realization" in which a combination of upper level elements is realized as a single lower level element, and "anataxis" in which two upper level elements occur in one order while their lower level realizations occur in the opposite order (cf. Lockwood 1972:27-29).

The term "sememe" is potentially confusing since it is used to refer both to class concepts and specific concepts.
The following diagram shows what the term sememe refers to.

A (class) sememe such as 'cat' consists of (cf. AND in the diagram) propositional and predicative relationships with other (class) sememes. These other sememes and their relations to 'cat' are called the "intension" of (or said to be true of) the sememe 'cat'. The predicative relation is represented in the diagram by a single symbol such as 'feline' and 'pet'. 'Cat' is in an "is a" relationship to these sememes; that is, a 'cat' "is a" 'feline', or a 'cat' "is a" 'pet'. Other parts of the intension are propositional; that is, a 'cat' 'eats-meat', in which 'meat' is a (class) sememe, 'eat' is a (class) sememe, and together they form a proposition which is true of 'cat'.

A (specific) sememe, often called a "particular", an "instance", or a "member" of the (class) sememe is indicated in the diagram by the (class) sememe name or a proper name, together with a suffixed number (cf. cat-1, Plum-1, etc.). This provides a unique symbol to represent our concepts of
individuals that we know or know of; for example, I know 'Plum-1', 'Max-1', and the two cats who live next door ('cat-1' and 'cat-2'). Plum and Max are suffixed with a number to distinguish them from other Plums and Maxes I might know or meet. Only (specific) sememes have real-world referents; (class) sememes do not.

1.3 Previous work

The discussion in 1.1 states briefly the aims of cognitive linguistics as opposed to those of descriptive or generative linguistics (cf. page 4). Cognitive linguistics builds on the foundation of descriptive linguistics, but has a rather more ambitious goal. While the aim of descriptive linguistics is to describe the structure of texts, cognitive linguistics attempts to account for the underlying mental system that makes it possible to produce and understand those texts. It is concerned with the cognitive system of the individual who, though one would hope is typical of the speech community, is by no means "ideal" in the sense that generative linguists propose. In addition, while generative linguistics claims a competence model of the ideal individual, it largely ignores a very important aspect of language as a cognitive system, the ability to understand (cf. the label "generative"). Neither do generativists deal with the actual production of real texts since they identify that task as performance and therefore out of their purview. In fact, generativists are unconcerned about that which people actually do, produce and understand language, the major
concern of cognitive linguists.

Many linguists of different theoretical persuasions, such as Randolph Quirk and Sidney Greenbaum and M.A.K. Halliday, have worked on English. We rely on the results of their research as a basis for building a cognitive model. Examples of previous research on English in the cognitive framework are Ikegami's (1969) treatment of verbs of motion, Bennett's (1970, 1975) work on prepositions, and Johansson's (1976) account of modal auxiliaries. Each of these attempts has concentrated on one or another specific aspect of English, whereas the goal of the present work is to address many aspects of the language, a core grammar as it were, and, in addition, to do so in such a way that the grammar is readily expandable.
NOTES

Chapter 1

1 The term 'Cognitive Linguistics' is to be understood in the sense that it is described in 1.1 and 1.3.

2 'Item' here refers to the object which was seen as either 'arranged' or 'processed' by linguists of this period. (cf. Hockett's 'Two Models of Grammatical Description' for further explanation of the terms in quotes.)

3 De Saussure (1916. in 1966:104) states that '...language does not offer itself as a set of predetermined signs that need only be studied according to their meaning and arrangement ...' This statement may at first appear to be in direct conflict with my interpretation of de Saussure above, but upon further investigation, we find that he is referring only to the delimitation of the signifier in the spoken chain. '... it is a confused mass, and only attentiveness and familiarization will reveal its particular elements. The unit has no special phonic character, and the only definition that we can give it is this: it is a slice of sound which to the exclusion of everything that precedes and follows it in the spoken chain is the signifier of a certain concept.'

Also, we must keep in mind that de Saussure is fixing signs for a single speech community only, not for languages in general. 'If words stood for pre-existing concepts, they would all have exact equivalents in meaning from one language to the next; but this is not true.' (de Saussure 1916. in 1966:116)

4 De Saussure does, of course, address language change over time (cf. de Saussure 1916. in 1966:74-8).

5 The following discussion is based on Lamb's 1985 oral address to LASSO and its companion paper 'On Defining the Language', presented at the 1986 meeting of LACUS.

6 We do this sort of thing all the time. Take, for instance, the concepts of 'land' and 'sea'. They are very real to us, yet can you demarcate exactly the boundary between them? Of course not, but we do not become disoriented because our concepts, which appear so concrete, are often illusory. Humans draw boundaries and categorize
to function and make sense of things.

7 This term is one used by Lamb in personal communication. Cf. also the Prague School's interpretation of the sign.

8 The discussion of sememe is based on class notes from J. Copeland at Rice University, 1983-1987.
CHAPTER 2
SYSTEM OVERVIEW

2.0 General Remarks

A sketch of Lamb's schema for a system which accommodates both the linguistic information and its use and enhancement in encoding and decoding is the topic of this discussion. A few suggestions of applications for such a system are presented in the last section (cf. 2.5).

Lamb has proposed a set of tools that is flexible in the important respect that it may be adapted by the system author (as distinct from the system user) as a strictly stratified system, a very integrated, partly stratified system, or any type in between.

This author has chosen to adopt a fairly strictly stratified version of the system which Lamb proposes, and that is the one elaborated in this and subsequent chapters. The reader should remain aware, however, that there is an alternative system architecture, and a person authoring a system using it would codify the linguistic information in a way different from that presented here. Lamb, in fact, prefers the alternative organization to the one being treated in this thesis.

A distinction is made between relations and processes. The linguistic information is relational, while the methods to modify that information and to encode and decode texts using that information are processes.
2.1 Basic system design

The version adopted here uses a series of dictionaries to house the linguistic information present in the system (cf. Chapters 3, 4, and 5 for the linguistic information). These dictionaries are intermediate to the levels of linguistic structure, and represented by the boxes in the graphic layout of this schema given below.

```
         SEMEMIC
            +----+
           /    |
          /     |
         +---------+
             |      |
             | SEMICON |
             +---------+

         LEXEMIC
            +----+
           /    |
          /     |
         +---------+
             |      |
             | LEXICON |
             +---------+

         MORPHEMIC
            +----+
           /    |
          /     |
         +---------+
             |      |
             | MORPHICON |
             +---------+

         GRAPHEMIC
```

The graphemic, not the phonemic, level is shown in the figure. The reason is three-fold: (1) graphemes form an alternate medium of expression to phonemes for language, (2) the description at the lowest level is simplified tremendously by using graphemes rather than phonemes, and (3) since the present treatment is a model for the computer, the input into the system is via the keyboard and, therefore, graphemic.

The names of the dictionaries are patterned after the term "lexicon" meaning dictionary. -icon is suffixed to the roots, morph-, lex-, and sem-, taken from the higher of the two levels of structure being mediated; thus, Morph-icon
intervenes between the graphemic and morphemic levels of linguistic structure, and so on. The term Lexicon is, therefore, to be understood, not in its usual, general sense, but in the specific sense to be derived from its relationship to the lexemic level of linguistic structure (cf. 1.2).

Within the dictionaries are entries which contain the linguistic information. Each entry, also referred to as a rule or statement, is written in a semiotic format which represents a sign or nection (cf. 1.1). The format of the entries is discussed in the following section (cf. 2.2).

While the dictionary entries may be referred to as 'rules', the reader is asked bear in mind that these rules differ significantly from rewrite rules. First, rewrite rules specify an operation; dictionary entries specify a relation (a sign relation). Second, the write operation in rewrite rules is unidirectional (indicated by the direction of the arrow in the rule); dictionary entries, with no built-in operation, can be interpreted in either direction. And third, rewrite rules change the representation from that which is on the left to that which is on the right (at the tip) of the arrow. Dictionary entries do not change or rewrite the representation. They simply state a relation.

Each dictionary has within it entries that could be characterized as either lower level or higher level. No physical separation of these types of rules is intended, however, as it is for the three main dictionaries (which are
in separate files). For example, in the Morphicon, entries for (morphemic) words (cf. 3.1-3.8) may be conceived of as being at a level lower than entries concerned with suffixes that can occur on those words (cf. 3.1 and 3.8). In the Lexicon, entries which account for idioms (cf. 4.1) are lower than statements that accommodate general phrase structure (cf. 4.2-4.19), since idioms are functionally closer to words than phrases. And in the Semicon, rules for semantic relations and clause structure (cf. 5.2-5.7) may be viewed as lower than the entries for meaning (cf. 5.1) and decoded, stored propositions (cf. 6.2). Some practitioners of stratificational linguistics have even proposed completely separate levels for the types of linguistic information that the present treatment recognizes as being on the lower and higher levels of the Semicon, calling clause structure the sememic level and meaning the hypersememic or gnostemic level. Lamb even suggests in his architectural schema (cf. fn. 1) a Gnosticon, placed above the sememic level.

The more stratified system design that the present approach employs is reflected in the dictionary entries. Stratification requires that the linguistic information be characterized in several, related entries (at the different levels of analyses) rather than in a single entry (as Lamb's alternate design with a master dictionary entails (cf. fn. 1)). For example, at the lowest level in the most stratified approach, the only information about the morpheme
'cat' which is present is that it is a noun. At a higher level, 'cat' is found to be a count noun, and at a still higher level, to have the sememic property 'animal'. All of these types of information are important from a global standpoint and are therefore present at some level in the system. From the local standpoint, however, that is, from local meaning to local expression, only the type of information needed for local co-occurrence restrictions is included in an entry. Entries are therefore particular to the level at which they occur.

The stratified codification of linguistic information, as illustrated by the above example for 'cat', defines one end of a continuum which accommodates degrees of strictness of stratification. The present treatment gravitates towards this strict end of the scale, but must be recognized as having at least one prominent digression from it. In the Morphicon, some lexotactic specifications are given for form classes like noun and verb. Rather than simply characterizing a morpheme such as 'cat' as a noun, an appropriate categorization based on morphological co-occurrence restrictions, the present approach also differentiates (lexotactic) sub-classes of the form classes and states them in the Morphicon. 'Cat' is therefore categorized as a count noun. Similarly, a verb such as 'eat' is not classed simply as a verb, but as a transitive verb, and so on. Other than this one deviation, however, we have tried to adhere to a model of as strict stratification as possible.
In the Morphicon, only the most productive derivational and inflectional morphemes are dealt with separately from the stems with which they co-occur. Forms such as 'took', 'taken', 'am', 'is', 'men', and 'rarely' simply get their own entries (cf. Chapter 3).

2.2 Notation for representing linguistic information

Dictionary entries conform to certain conventions. Each entry has a left side and a right side, which are separated by a backslash (\). The left side is the (local) expression; the right side is the (local) content. The general format of an entry is:

expression \ content

As Lamb points out, the slant of the backslash produces '...an upper side and a lower side to indicate which side is which, in accordance with the diagramming convention of putting content above expression.' (cf.fn.3)

The reader may simply interpret the backslash as 'is a realization of' when reading the rule from left to right (cf.1.2).

To represent the elements (above the graphemic level) in each entry, we use words and abbreviations of words as symbols, due to their mnemonic value. More abstract symbols could have been used, but readability would have been forfeited. The important thing to remember is that the symbol, whatever it is, represents a single unit. For example, the word 'cat' is the symbol used to represent
either a morpheme, a lexeme, or a sememe, according to which
dictionary it is in and which tactic code is associated with
it. Therefore, if 'cat' appears on the right side of a rule
in the Morphicon (the left side is the graphemic
representation), it may be understood to be a symbol for the
morpheme. If it appears on the right side of a rule in the
Lexicon, it is construed as a lexeme. On the right side of
a rule in the Semicon, it is to be taken as a symbol of the
sememe.

Sometimes the symbol is an abbreviation of a word.
'Pl' denotes plural, 'Sg' means singular, 'Pres' means
present, and so forth. Each time one of these abbreviations
is encountered for the first time in the rules in Chapters
3, 4, and 5, it is defined for the reader.

As with the elements, tactic labels may be symbolized
either by a word or an abbreviation. Some of these labels
are Mass, Head, Focus, N (noun), NP (noun phrase), Adj (ad-
jective), and so on. They too are defined in the appro-
priate chapter.

Tactic labels may have as many suffixes as the author
of the system desires and deems necessary to make the symbol
explicit (i.e., to make the symbol unique for the computer
or mnemonically useful to the human). Such tactic suffixes,
if present, are separated by periods. Some example tactic
labels, with their values given in parentheses, are:

Nmr (number)
Tns (tense)
N.Cnt (count noun)
Det.Indef.Cnt (indefinite count determiner)
Adj.Age (age adjective)
V.T (transitive verb)

The label that comes first (or stands alone, cf. 'Tns' and 'Nmr') denotes the main category. The suffixes on the label identify sub-categories that further specify the main category.

More than one word or abbreviation may be used to symbolize elements as well. When it is, a hyphen joins the parts. For example, the concord element on verbs, inflected for third person singular, is represented as '3-Sg'. Complex lexemes take advantage of this convention as well, for instance 'look-in-on' and 'foot-loose'. These hyphenated forms for elements are to be understood as simple symbols, however. Their apparent complexity is a product of the notation, not of the element itself.

All of the abbreviations, whether for the elements or tactic codes, used in this thesis can be found in Appendix B.

Elements and tactic codes are separated from each other by a colon (:) . Thus, there are portions of entries like the following:

N.Cnt:cat (the count noun 'cat')
V.T:eat (the transitive verb 'eat')
Adj.Age:old (the age adjective 'old')
Tns:Pres (the tense 'present')

Symbols for elements usually begin with a small letter unless they are abbreviated, in which case they begin with a capital letter, as shown immediately above (cf. 'Pres'). There are a few exceptions to that general rule: 'Past' and
proport nouns such as 'Christmas', 'Houston', etc.

Tactic codes, including suffixes, always begin with a capital letter (cf. Adj and Age, above). Some tactic codes have other capital letters also, even if they do not have suffixes.

NP (noun phrase)
PartP (participial phrase)
PP (prepositional phrase)
RelvCl (relative clause)

Not every element has a tactic code coupled with it. These elements usually constitute a class by themselves. In the present grammar, such elements are 'm', the objective case marker of pronouns; 'not', the negative; 'en', the past participle marker; 'ing', the present participle marker; and so on. Graphemes do not have tactic labels either.

A grapheme is an element which is represented by a letter of the alphabet or other symbol found on the keyboard of a typewriter. For example, the grapheme 'a' is represented by the symbol 'a'. The expression side of a rule in the Morphicon shows graphemic strings. A graphemic string is a succession of graphemes; for example, the graphemic string that represents the morpheme 'cat' is "c a t". A space separates each element from the others. Thus, the full entry for 'cat' in the Morphicon is:

c a t \ N.Cnt:cat

In many rules, it is impossible to state an exact element or tactic sub-type because the possibilities are too great to enumerate. Such a rule is called a general rule, and to indicate the unspecified element or tactic sub-
category, a variable is used. Variables in entries are limited by the designer of the notation system (Lamb) to X, Y, or X or Y followed by an integer (e.g. X,Y,X1,Y1,X2,Y2, etc.). Variables specified in a rule are particular to that rule; that is, if X is used in two different rules, it does not indicate the same item, except coincidentally. The same variable occurring in the same rule, however, does indicate the same item. X in the following portion of a rule means an element which is a count noun.

N.Cnt:X

In some cases, the co-occurrence restrictions that an entry is concerned with do not require any specific sub-type of, for example, noun; that is, the rule is a general rule applicable to any type of noun. Two variables are used to indicated the unspecified parts.

N.X:Y

where X is the noun (N) sub-type, and Y is the element (morpheme).

A variable is never used to indicate an element and its tactic label, for example, X cannot be used for N.Cnt:cat. A form such as N.Cnt:cat is represented as X:Y, where X represents the tactic label and Y the element.

When, in the course of decoding or encoding, a value is given to a variable in a general rule, that process is known as "instantiation". If the variables in N.X:Y are given the values 'count' and 'cat', respectively, then we say that X is instantiated to 'Cnt', and Y is instantiated to 'cat'.
Parentheses figure importantly in grouping certain items together. Those grouped items may, in turn, have a tactic label. For example, a nominal (Nom) consists of a head (Head) and modifier (Modif) (disregarding for the moment that the modifier is optional). The notation which expresses that statement is:

Nom: (Modif: X Head: Y)

Nominal refers to everything inside of the parentheses.

Embedded sets of parentheses are common in rules. Within the nominal, for instance, parts of the modifier or head might be stipulated.

Nom: (Modif: (Adj.Age: X) Head: (N.Cnt: Y Nmr: X1))

The lines and arrows underneath this portion of a rule match up the opening and closing parentheses of each pair.

The X and Y in this rule are not co-terminous with those in the above stated version of the rule (cf. Nom: (Modif: X Head: Y)). X, in the former, refers to that which (Adj.Age: X) refers to in the latter; and similarly, Y, in the former, refers to that which (N.Cnt: Y Nmr: X1) refers to in the latter. A variable can, therefore, stand for a whole string of things. That whole string, however, must be a single unit, a constituent group. A variable is never used to mean several things that are not grouped. For example, X never represents a string like this: N.Cnt: cat Nmr: Sg. Only when those constituents are grouped, as in Head: (N.Cnt: cat Nmr: Sg), can they be referred to by a
variable, as in Head:X.

Square brackets ([]) signify optionality. For example, a modifier is an optional constituent of the nominal.

Nom:([Modif:X] Head:Y)

Sets of square brackets can be embedded like sets of parentheses.

[Fut [Perf]]

The lines and arrows show the pairs of square brackets. 'Fut' (future) and 'Perf' (perfect) are optional constituents, signified by the outer set of brackets. They may both be present or both be absent. The inner set of brackets makes 'Perf' optional with respect to 'Fut'. 'Fut' may be present with or without 'Perf', but not vice versa (cf. 4.18). The possible combinations stipulated by this notation are:

Ø (nothing)
Fut
Fut Perf

A comma denotes an OR relationship. The scope of the OR relationship is delimited by a set of parentheses. For example, 'orange' may be a count noun OR an adjective of color in the following rule.

(N.Cnt, Adj.Col):orange

The possible choices provided by this notation are:

N.Cnt:orange
Adj.Col:orange

Such an interpretation implies that the comma has precedence
over the period.

More than one element may be specified as part of a choice. Whatever there is, from the opening parenthesis to the comma, between commas, or between a comma and the closing parenthesis, is included in the choice.

\[(en \ \text{Adv.X:Y} \ V.X1:Y1, \ \text{Adv.X:Y} \ NP:X2)\]

provides these choices:
en \ \text{Adv.X:Y} \ V.X1:Y1
\text{Adv.X:Y} \ NP:X2

OR relationships may be embedded.

\[(\text{Aux:}(\text{have,do}), \ V.Be:be)\]

provides these choices:

\text{Aux:} have
\text{Aux:} do
\text{V.Be:} be

Square brackets also delimit OR relationships. The following elements are not hierarchically optional as 'Fut' and 'Perf' are, shown above; rather they participate in a simple OR relationship with one another.

\[[\text{Adv.X:Y}, \ \text{Intns:X1}]\]

The alternatives provided by this notation are:

\(\emptyset \ (\text{nothing})\)
\text{Adv.X:Y}
\text{Intns:X1}

The asterisk signifies repeatability: \(X^* = X, XX, XXX,\) etc. An asterisk after an element or constituent group enclosed in parentheses denotes the same possibilities, so \((X \ Y)^* = X \ Y, X \ Y \ X \ Y, X \ Y \ X \ Y \ X \ Y,\) etc. If the asterisk follows an element or constituent group delimited by square brackets, then zero (nothing) is also a possible alterna-
tive, for instance \([X \ Y]^{*} = \emptyset, X \ Y, X \ Y \ X \ Y,\) etc.

When elements or constituents can occur in any order with respect to one another, the anataxis (cf.1.2) is indicated by a tilde (\(~\)) in front of a constituent group which is delimited by parentheses (cf.3.3, 3.6, 4.5, 4.12, and 5.6).

\(~(X \ Y)\)

provides for the following interpretations:

\[X \ Y\]
\[Y \ X\]

The function of the tilde can be remembered by its shape which resembles an editor's mark for transposition.

This notation can be used in conjunction with other notation to allow many options, for example (cf.4.19):

\(~([\text{Adv.} X: Y] [\text{PP}: X])^{*}\)

Both the adverb (Adv) and prepositional phrase (PP) are optional (indicated by []), and the whole thing is repeatable (indicated by the asterisk), but the adverb and prepositional phrase may occur in any order with respect to one another as, for example, in 'I hit him in the rear end there by the license plate'. The underlined portion consists of a prepositional phrase ('in the rear end'), followed by a locative adverb ('there'), followed by another prepositional phrase ('by the license plate'), any one or all of which are optional constituents.

In the Morphicon, on the expression side of a rule, a hyphen not immediately preceded by another symbol (i.e., after a space) is used to indicate a suffix (cf.3.1, 3.8).
Two examples with a hyphen are:

N.X:Y -s
V.Be:be -n 't

The grapheme 's' in the graphemic string "c a t s" illustrates the first example; the graphemes 'n', '''', and 't' in the graphemic string "a r e n ' t" illustrate the second example.

To identify a grapheme which is the final part of the realization of a morpheme, a hyphen with that grapheme symbol is enclosed in a pair of angle brackets (<>), with no space intervening between it and the variable denoting the morpheme (Y).

N.X:Y<-s>
The 's' is the final character of Y. 'Christmas' and 'guess' are two such morphemes whose graphemic realizations end in 's'.

Spelling conventions are such that alternative representations are often used under different circumstances. For example, the symbol for 'story' (represented by Y in the rule below), though ordinarily ending in 'y', ends in 'i' before the graphemic string "-e s". The alternatives in such an environment are separated by a slash (/) (cf.3.1).

N.X:Y<-y/i> -e s

Angle brackets found in the Semicon indicate a presupposed participant. In the following rule, X presupposes Y, but not vice versa.

X<Y>
If $X = 'sleep'$, for example, then $Y = 'animal'$ because the event 'sleep' presupposes one participant which is some type of 'animal' (the category animal includes humans (cf.5.1)). (Metaphorical usages of 'sleep' are not considered here.) This presupposition is part of the meaning of 'sleep' (cf. 5.1).

'Ø' is the symbol for zero. It is used primarily in situations of "empty" and "zero" realization (cf.1.2).

\[
\text{not Aux:do} \setminus \text{Neg Aux:Ø}
\]

If no auxiliary (stipulated by Aux:Ø) co-occurs with negative (Neg), then 'do' is realized. Ordinarily, two things on different sides of a rule are co-referential only if they have the same symbol, as in: $X \setminus X$. The above rule shows that co-reference is specified by different symbols (in this case Ø and 'do') because they are otherwise shown to be related by their immediate context (in this case by the tactic code "Aux").

Two dots (..) signify an unspecified string of indefinite length.

$X .. Y$

This notation indicates that $X$ is followed by $Y$ with some intervening string of undetermined length (which may be zero). The way in which the system determines the length of the unspecified string during processing is discussed in 2.4.

The IF statement is a part of many rules. Its purpose is to specify exactly the range of possible instantiations
for a variable. It may be written in either a positive or negative form. The following rule fragment shows a positive IF statement.

\[ \text{NP:} (\text{Head:X Nmr:Y}) \quad \text{IF} \quad X = (I, you) \]

where 'I' and 'you' are the only two possible values of X.

An example of a negative IF statement is:

\[ \text{NP:} (\text{Head:X Nmr:Y}) \quad \text{IF} \quad X \not= (I, you) \]

X is limited to any element other than 'I' and 'you'.

In the Semicon, properties constituting intensional information (cf.1.2) about a sememe are joined by a plus sign (+) (cf.5.1).

\[ \text{sememe} \backslash a + b + c \]

To stipulate that more information is present than is specified, 'etc.' is used.

\[ \text{sememe} \backslash a + b + c + \text{etc.} \]

For semantic relations, properties of sememes must often be stated. These properties appear between a pair of dollar signs ($). The dollar sign acts as a mnemonic device because the S in it stands for sememic (property).

\[ X \ Y \ X \ Y \ \$X1$ \]

A tactic combination of X followed by Y is acceptable if Y has the sememic property X1. That which is enclosed in dollar signs refers to the immediately preceding element. For example: at ten-o'clock \at ten-o'clock $\text{temporal}$

The preposition 'at' can only co-occur with an object ('ten-o'clock') that is temporal.

All of the symbols discussed in this section are
encountered again in Chapters 3, 4, and 5, where the actual rules of the grammar are stated. They are elaborated on further there. The reader is reminded that abbreviations for words which are used as symbols appear with their full forms in Appendix B.

2.3 The dynamic aspect

Furnishing a basic grammar such as the one elaborated in Chapters 3, 4, and 5 of this thesis is (1) our primary goal, and (2) the first step in building a functional natural language processing system. Of course, any system which is to handle information, be it a human system or an artificial one, must be able to continually modify its existing information base. We therefore present an informal sketch of some methods for modifying the system we have formally elaborated. This modification, also known as system editing or updating, is a process of adding new dictionary entries or refining existing ones.

Information is usually added when a new lexical item is encountered. This operation requires at least two dictionary entries, one to the Morphicon for its initial analysis into morphemes and one to the Semicon to state its meaning. Information to be added can not be gained directly simply by asking the (naive) user for it, however. More sophisticated methods, accessing the preexisting store of linguistic information, must be employed. Section 6.1 is a discussion of such techniques.
If a dictionary entry accurately reflects the information, why should it ever be necessary to further refine it? Because information may be incorrect or incomplete. For example, if someone tells us that veal is the meat of deer, and later we find that it is the meat of a young calf, we want to be able to amend the connection made between veal and deer, and make it for veal and calf. This essentially means refining the content side of the Semicon entry for 'veal'.

Information is always incomplete. When we learn a new word, for example, the meaning that is stored reflects its use in the particular context in which it was heard (or read). As the word is encountered again and again in other contexts, however, we must refine our notion of its meaning to accommodate all of its uses.

Not only words are added or refined, however. Syntactic constructions are subject to similar treatment as well. Updates to the Lexicon are also therefore possible.

Lamb states that the above types of addition and refinement are simply language acquisition (Lamb, forthcoming). Thus a child or an adult is always acquiring language.

Lamb differentiates this type of structure building from that which he states occurs in the interpretation of every phrase decoded (Lamb, forthcoming). Whereas the former may add or refine structure at any level, the latter does so only at the conceptual level. These decoded
messages comprise the individual's storehouse of knowledge about the events that have taken place in his (personal or vicarious) experience. Section 6.2 discusses techniques for identifying and storing decoded information.

2.4 The procedure for encoding and decoding

Lamb (1985:150) states that encoding and decoding are processes which are distinct from the linguistic information housed in entries in the dictionaries. Encoding and decoding use the information in these dictionaries, applying it to some input. Essentially, the difference is one of data (the dictionary entries) versus process (encoding and decoding).

```
/\   SEMEMIC
   |   +-------------+
   |     | Semicon |   E
   |     +-------------+
   |     | LEXEMIC     |   N
   |     +-------------+
   |     | Lexicon |   C
   |     +-------------+
   |     | I          |   O
   |     +-------------+
   |     | MORPHEMIC   |   D
   |     +-------------+
   |     | Morphicon  |   I
   |     +-------------+
   |     | GRAPHEMIC   |
```

Lamb (1985:151-2) has proposed that processing can be viewed as simply a multi-layered dictionary look-up procedure which attempts to match some (contiguous) portion of the input string to one of the entries in the dictionary. If decoding, the attempted match is to the left (expression) side of some dictionary entry. If encoding, the match is to the right (content) side of an entry (cf. 2.2). The routine
that is implemented in the event of a non-match is explained in 6.1 (cf. 2.3).

Aside from their organization in separate dictionary-like modules, rules are not ordered with respect to one another. The way in which the rules are written allows matching to occur only when the input has been analyzed to a sufficient degree.

How is a match accomplished? The input string for decoding comes in via the keyboard and has spaces or other punctuation between words. These spaces, commas, and so forth act as delimiters, and each letter in the word between the spaces (etc.) is analyzed as a single unit, a grapheme. Thus, the internal representation of the following input consists of three graphemic strings.

Input: I am hungry.
Internal representation: I
a m
h u n g r y

The Morphicon is the dictionary with graphemic information in it, so the system attempts to match each string of graphemes in the internal representation of 'I am hungry' to the left sides of entries in the Morphicon. When a match is made, the right side of the rule is given as the higher level representation, which is used in attempts at matching at the next level. The successive matching processes continue until there are no more rules which can apply, and the original input has been decoded to the clause level. What is done with the decoded clause is discussed briefly below
and in more detail in 6.2.

Decoding generally proceeds by using the entries in first the Morphicon, then the Lexicon, and then the Semicon. The encoding process takes the opposite route through the dictionaries, using first the Semicon, then the Lexicon, then the Morphicon. The major exception to this sequential use of the dictionaries occurs when there is a "rank-shift" of a phrase or clause. "Rank-shift" is Halliday's term for "...the transfer of a given unit to a lower ..." level (1961:251). It is a reduction of a higher level unit of structure such that it functions as a lower level unit. For example, in 'The woman who lives next door's cat', 'the woman who lives next door' is a (complex) noun phrase (on the lexemic level) that has been rank-shifted to the status of noun so that it can co-occur with the inflectional possessive morpheme '-s' (on the morphemic level). Because the noun phrase has been reanalyzed as a lower level unit (as a noun), it may enter into phrase structure constructions in its new capacity, and that is what this noun phrase cum possessive noun has done. It is acting as a modifier to the noun 'cat', and together they form a(nother) noun phrase.

Processing thus begins with Morphicon rules, next uses Lexicon rules, goes back to Morphicon rules, then goes again to Lexicon rules, and afterwards proceeds in its usual manner. The processing procedure is alerted to implement this "dictionary hopping" by certain elements such as 'Poss' (possessive) or 'Relv' (relative (pronoun)), the element that marks
the beginning of a relative clause (cf.3.3). We return to 'Relv' again below.

Processing is performed in a parallel manner and any rule that can apply, will apply. Branching, the setting up of alternative paths to follow, occurs when more than one rule matches the current (local) analysis, thus producing two or more (local) analyses. Usually, it is desirable to have only one terminal decoding of an input (with the obvious exception of puns), so the rules are written in such a way that processing along undesirable paths dies out, hopefully at the level where the ambiguity would be resolved by the typical speaker/hearer. For example, in decoding 'a farm', 'farm' will initially be analyzed as both a noun and a verb, creating a branching in the processing. There is an entry which incorporates 'a' and noun 'farm' into a noun phrase, but no construction allows for a combination of 'a' followed by verb 'farm'. Thus, the verb path of the branching dies out at the lexemic (phrase structure) level.

When a clause is decoded, something must be done with it. The specific action the processor takes depends on the type of clause decoded. A declarative clause should prompt an ADD command, because the use of a declarative is a request to the interlocutor to add information to his system (cf.fn.8,p.7-2). An interrogative is a request to the interlocutor to find information that he already possesses and report it.

'Relv' (cf.3.3) is an element which cues the system to
begin processing a clause at the point at which it occurs in the input string. In Section 2.2, we said the two dot notation (..) in an entry signifies an unspecified string of indefinite length. These two symbols ('Relv' and ..) function together to process the relative pronoun at the proper place in the string. Although, at the surface level, relative pronouns always come at the beginning of the relative clause, the traditional slot for grammatical subject in English, they may function as either the subject or object, for example:

( the one ) which hit me (which = subject)  
( the one ) which I hit    (which = object)  

In the former example, the relative pronoun 'which' is the subject of the relative clause, telling what did the hitting; in the latter example, it is the object, telling what got hit. The entry in the Morphicon for 'which' is (cf.3.3):

w h i c h .. \ Relv .. Pro:which

For the first example, the two dots from the above rule are instantiated to zero since Pro:which, acting as subject, occurs immediately after 'Relv' which demarcates the beginning of the relative clause:

w h i c h \ Relv Pro:which

For the second example, the two dots are instantiated to 'I hit' since, in unmarked word order, they occur between 'Relv' (marking the start of the relative clause) and Pro:which that is acting as object of the verb 'hit':

w h i c h (I hit) \ Relv (I hit) Pro:which
(The portion in parentheses is what the two dot notation covers.) In essence, the two dots (in combination with an element like 'Relv') put the element following it "on hold" until it is needed to complete some construction (in this case, a phrase structure construction which requires an object to follow the transitive verb).

'Relv' is not the only element that can cue such an action. 'Intr' (interrogative) acts in a similar manner (cf.3.3,3.6). The traces of decodings of relative clauses and interrogatives in Chapter 7 illustrate this point further (cf.7.2 and 7.3).

The two dot notation, when not accompanied by an element like 'Relv' for example, simply denotes anything in the input intervening between those portions specified.

NP:((..Head:(X Y)))

In this example, the two dots mean anything in the noun phrase that occurs before the Head: the determiner, numerals, modifiers, etc.

2.5 System applications

Applications are not part of the system outlined in the preceding sections of this chapter. Rather, an application is a possible use of (some or all of the capabilities of) the system to perform a specific task. In this section, we outline briefly the parts of the system which two possible applications might use.

The two applications of a system such as the one model-
led in this thesis are one, a translation device, and two, a knowledge base for information storage and retrieval.

Machine translation can be developed to perform with varying degrees of sophistication. At one end of the scale, there is "rote" translation which essentially accommodates sentence-length texts and requires a fair bit of human post-editing. At the other end of the scale, still very theoretical, is a translation device which simulates the human in the sense that it accommodates larger-than-single-sentence-length texts, remembers what it has translated, and needs little human post-editing.

The former type, a rote translation device, is the application which is discussed here and again in Chapters 6 and 7. Rote means 'a fixed, mechanical way of doing something, ... without understanding or thought' (Webster's 1980:1238). A translation device that works on this principle takes some input sentence, decodes it to the clause structure level, then passes it to a translation dictionary without having "understood" it. This statement needs some explanation.

In 6.2, understanding is distinguished from decoding. As Lamb (forthcoming) states, 'Understanding of a text consists of relating the results of its decoding to the already present information.' Some further actions, such as adding, verifying, or retrieving information, should, therefore, be taken, after the decoding to the clause level has been accomplished. In a rote translation application, these
further steps are omitted, and the decoded clause is passed immediately to the translation dictionary.

A translation dictionary serves as an interface between two languages. It should relate the symbols used to represent the sememes and semotactics in the source language to those used in the target language. Which language is source and which is target depends on the direction of translation. The entries in a "transfer" dictionary can be written as bidirectional. We do not include an in-depth discussion of the transfer dictionary since it is not part of the system described in this dissertation.

A knowledge base is another possible application of our system. A knowledge base is a store of information to which more information may be added or from which information may be retrieved. This application is sometimes called a question-answering system and is usually further differentiated into sub-application areas, each of which controls some specific domain of information. Some examples of such domains are facts about baseball, medical records of patients, or information on general linguistics.

All of the aspects of the system outlined in this chapter, the linguistic information and the dynamic capabilities, are necessary for a question-answering application. To build the knowledge base about a specific domain, information must be stored. That requires not only the decoding procedures and the current linguistic information, but also the dynamic aspects of the system which add
linguistic information like new words and their meanings, and conceptual representations which indicate understanding (and thus allow retrieval). Information retrieval also depends on the decoding procedure and the current linguistic information to decipher what the user wants. It also depends on a dynamic search and seizure operation and on an encoding procedure to get and return an answer to the user.
NOTES

Chapter 2

1 Abridged version of Figure 4-1 in 'Notes on Dynamic Grammar: Section 4. The Linguistic Description as a Dictionary-Like System', Sept.87. Lamb also gives in this paper an alternative system architecture in which a master dictionary houses the linguistic information for general cases, with separate supplement dictionaries for irregular word, phrase, and clause structure.

2 Lamb first proposed these terms in 1966 in 'Outline of Stratificational Grammar'.

3 From 'Notes on Dynamic Grammar: Section 5. Notation for Language Description', Nov.87, p.5-1.


5 Lamb's alternative, more integrated approach has all the information needed globally given at once in the master dictionary. For example, the entry for 'cat' might include noun and animal, with a separate inference rule that all animals are concepts that are realized lexemically as count nouns.

6 The general format of the entry has been suggested by Lamb (cf.fn.3). Some of the specific operators and other non-linguistic symbols emerged from my meetings with Dr. Lamb as the need for each one arose. The linguistic symbols (for elements and their tactic labels) were, for the most part, left to my discretion and largely conform to common terms in language and linguistics.


8 Ideas for this discussion are generally based on Lamb's 'Notes on Dynamic Grammar: Section 7. Techniques for System Modification', Dec.87, p.7-1 - 7-3; and notes from Lamb's classes and colloquia at Rice Univ. from 1983 - present. Of course, as elsewhere, any divergences from his ideas are my
own responsibility.

9  "Edit" and "update" are both used in their data processing senses meaning 'add, change, or delete'.

10  The actual internal representation in the computer is a string of ASCII codes, but this is immaterial for the present discussion.

11  The particular type of search procedure, binary or sequential, is immaterial to the present discussion which is outlining a general procedure for decoding and encoding.

12  Cf. the hold register approach in ATNs in Winograd (1983:232ff.).
CHAPTER 3
THE MORPHICON

3.0 General Remarks

The grammar presented in this and the following two chapters is not intended to be complete. The discussion in Chapter 1 concluded that language is a mutable system that is constantly being updated, making a fixed, well-circumscribed grammar a theoretical impossibility. Furthermore, no claim is being made that all possibilities for tactic categories, or their members, are being covered, even superficially. What is presented, however, is a grammar that can accommodate a great many English utterances as it presently stands or with only minor vocabulary enhancements. Since this is the first attempt to codify English into this new semiotic format, this work is not only a grammar of the language, but is also an example of one way to encapsulate the linguistic information of a language in this format.

The grammar itself is based to a large extent on the descriptions of English by several linguists, who will be referenced where appropriate, but what has been attempted here is not just a description of texts, but the construction of a model which will account for those texts (cf.1.3).

Accompanying the presentation of the grammar are a few semantic explanations for the tactic distinctions that have been observed. Although the tactic distinctions rest on
distributional criteria, the reasons for such distributional facts can most often be found in semantic factors. Some of these factors are included in the discussion.

The description of the grammar is presented in three chapters. Chapter 3 covers the entries in the Morphicon, Chapter 4 the Lexicon, and Chapter 5 the Semicon. The reader is reminded that these three separate dictionaries function together as a whole, and the particular entries in each involve connections to entries elsewhere in the system. To fully understand the usefulness of any rule, it must be considered in its larger context, as part of the system. Since we can not discuss everything at once, however, we include many cross-references to guide the reader to related sections for further details.

As discussed in Chapter 2, the approach we take in the treatment of linguistic structure is one of stratification. This requires that all information about elements not be stated at once, but be introduced at the level where it is used. The reader will notice that strict stratification is not observed, however, in the Morphicon where lexotactic codes (such as "count" noun) are assigned to the morphemes, rather than simply assigning them to natural classes based on their morphological co-occurrence restrictions (cf.2.1).

The rules are written in such a way that they may be used equally well in either the decoding or encoding process. The reader is reminded, however, that the linguistic information given in this semiotic format is separate
from the processes of encoding and decoding.

All entries presently in the grammar (cf. Appendix A) are numbered with a four-digit code. Any number beginning with '1' can be found in the Morphicon, with '2' in the Lexicon, and with '3' in the Semicon. These numbers have no further purpose in the grammar and are present only for reference.

For clarity of statement, only relevant portions of some entries are given. Such entries may be seen in their entirety in Appendix A, at the location indicated by the four-digit number in parentheses which follows the entry in the text.

Every entry contains a backslash whose significance is explained completely in 2.2. The reader is simply reminded to interpret the backslash, when reading from left to right, as 'is a realization of'.

We begin the discussion of the grammar at the level of the Morphicon because, as Lamb has suggested, a grammar for the hearer/reader is preferred over a grammar for the producer. This preference is based on the fact that children learn their language from hearing/reading, and adults also add new items of structure via these mediums.

The major part-of-speech categories that we are concerned with in our description are nouns, pronouns, numerals, determiners, adjectives, adverbs, verbs, and prepositions. These categories and their sub-classifications can be found in the Morphicon, the dictionary that
mediates the graphemic and morphemic levels of structure (cf. 2.1).

3.1 Nouns

The entries for nouns are divided into three major categories: proper nouns, count nouns, and mass nouns. The need to treat these elements under separate sub-classifications has of course to do with the restrictions applicable to them in their co-occurrence with other elements. The semantic explanation for the distinction between count and mass nouns is that things whose quantification is conceptualized as depending on their being contained by some other thing are termed mass, for example, 'two glasses of water' or 'a modicum of decorum'; whereas things whose quantification uses intrinsic conceptual boundaries are termed count, as in 'two cats' or 'an apology'. Some examples of count and mass nouns found in the Morphicon are as follows:

\[
\begin{align*}
\text{cat} & \text{ N.Cnt:cat} (1020) \\
\text{story} & \text{ N.Cnt:story} (1022) \\
\text{woman} & \text{ N.Cnt:woman} (1023) \\
\text{honey} & \text{ N.Mass:honey} (1050) \\
\text{work} & \text{ N.Mass:work} (1051) \\
\text{water} & \text{ N.Mass:water} (1054)
\end{align*}
\]

The left side of each of these rules shows a graphemic string. The right side shows the morpheme, specified at the right of the colon, and represented by a symbol consisting of one or more letters (cf. 2.2). The morphemes have either the tactic code count noun (N.Cnt) or mass noun (N.Mass).

While the count and mass distinction is fairly clearcut for most nouns, there are some nouns that easily belong to
either category. One of these is 'time'.

\texttt{time} \textbackslash \texttt{(N.Cnt, N.Mass):time (1081)}

There is a difference in meaning correlated with this difference in tactic classification. It can be seen in examples such as 'I've heard that many times', in which 'time' is classed as a count noun, or in 'I don't have time', in which it is a mass noun.

The third classification of nouns is for names of people, places, months, publications, and so forth. These are called proper nouns. Quantification of proper nouns or their combination with determiners does not usually occur since their reference is considered unique (Quirk and Greenbaum 1975:76). If the situation calls for it, however, as for example in, 'There are three Doug's in my ballet class', then proper nouns can be quantified or otherwise made less than unique (cf.4.5).

Some examples of proper nouns (N.Prop) from the Morphicon are:

\begin{verbatim}
R u d o l p h \ N.Prop:Rudolph (1001)
H o u s t o n \ N.Prop:Houston (1002)
F e b r u a r y \ N.Prop:February (1003)
M o n d a y \ N.Prop:Monday (1004)
C h r i s t m a s \ N.Prop:Christmas (1005)
\end{verbatim}

Since the morpheme at the right of the colon may be represented by any symbol desired, the symbol chosen begins with a capital letter so that it looks just like the graphemic representation and is therefore more readable.

As discussed in Chapter 2, only the most productive of the derivational and inflectional morphemes that might be
recognized in descriptive linguistics are treated here as separate morphemes (cf. 2.1). Where possibilities of recognizing non-productive ones are present, we prefer to put the whole inflected or derived word or stem in as a unit in the Morphicon. Some entries representative of this (portman-
teau) treatment are:

\[
\begin{align*}
\text{w o m e n \ N.Cnt:woman Nmr:Pl} & \quad (1024) \\
\text{m e n \ N.Cnt:man Nmr:Pl} & \quad (1026)
\end{align*}
\]

These particular forms are inflected for plural. The plurality is shown by the accompanying number (Nmr) morpheme plural (Pl).

Like 'women' and 'men' are 'children', 'oxen', 'data', 'criteria', 'alumni', 'geese' and so forth.

-s, or its alternant -e s, is the realization of the productive form of the plural morpheme in written English. Since it is so prevalent, it requires its own separate entries.

\[
\text{N.X:Y -s \ N.X:Y Nmr:Pl} \quad (1980)
\]

The grapheme 's' occurring as a suffix on the noun (N.X) Y is the realization of plural.

\[
\text{N.X:Y<-s, -z, -x, -ch, -sh, -y/i> -e s \ N.X:Y Nmr:Pl} \quad (1981)
\]

The string of graphemes "e s" occurring as a suffix on the noun Y is also the realization of plural, provided that Y as listed in the dictionary ends in s, z, x, ch, sh, or y (shown in angle brackets <<>); but if Y ends in y, then the graphemic representation has i rather than y before the suffix -e s. (This rule does not account for nouns ending
in y preceded by a vowel, for example, 'key:keys'.)

For example, the first of these two entries, (1980),
takes care of a graphemic string such as "m i t t e n s".
Although there is no entry in the Morphicon which has
"m i t t e n s" as its expression (left) side, there is this
entry (1980) for the suffix -s, and an entry for the count
noun 'mitten' (1031), so the linguistic information
necessary to accommodate such an input is present in the
system.

Graphemic strings such as "b o x e s" and
"s t o r i e s" are taken care of in a similar manner by
(1981). Although there are no entries per se in the
Morphicon for either string, both end in -es, which this
entry provides for. Both strings also have nouns, specified
by (1981), which are present in the Morphicon (cf.1022 and
1032) and which end in one of the specified ways (shown in
<> in the rule) - 'box' in x, and 'story' in y which is
represented as i when followed by -e s.

An element such as 'wives' has this productive form of
the plural morpheme, -s, but the entry for the singular form
is not 'wive' but 'wife' (1033). This alternation in the
stem is provided for in the same manner as for the irregular
plurals above. A separate entry is present in the Morphicon
for 'wives'.

\w i v e s \ N.Cnt:wife   Nmr:Pl \ (1034)

3.2 Determiners

The major division of determiners is into the sub-types
definite and indefinite. This classification of determiners approximates that described by Quirk and Greenbaum (1975:62, 66-69), but the reader is advised to refer to their discussion for their exact characterization.

Definite is a signal that some member of the class designated by the noun following the determiner is identifiable either from previous mention, shared world knowledge, or proximity. Indefinite indicates a message from the speaker to the hearer that the identity of the class member can not be established using any of the abovementioned ways, and therefore should simply be added for future reference. (Cf. 6.2 for a discussion of how this information gets added to the system.) For a discussion of degrees of identifiability, the reader is referred to Copeland and Davis (1980, 1981) and Davis and Copeland (1980).

Some entries in the Morphicon for definite and indefinite determiners are:

\[
\begin{align*}
\text{the} & \, \text{Det.Def:the} & \text{Nmr:(Sg,Pl)} & (1200) \\
\text{this} & \, \text{Det.Def:this} & \text{Nmr:Sg} & (1201) \\
\text{that} & \, \text{Det.Def:that} & \text{Nmr:Sg} & (1153) \\
\text{these} & \, \text{Det.Def:these} & \text{Nmr:Pl} & (1202) \\
\text{those} & \, \text{Det.Def:those} & \text{Nmr:Pl} & (1203) \\
\text{a} & \, \text{Det.Indef.Cnt:a} & \text{Nmr:Sg} & (1220) \\
\text{every} & \, \text{Det.Indef.Cnt:every} & \text{Nmr:Sg} & (1221) \\
\text{each} & \, \text{Det.Indef.Cnt:each} & \text{Nmr:Sg} & (1222) \\
\text{either} & \, \text{Det.Indef.Cnt:either} & \text{Nmr:Sg} & (1223) \\
\text{neither} & \, \text{Det.Indef.Cnt:neither} & \text{Nmr:Sg} & (1224) \\
\text{much} & \, \text{Det.Indef.Mass:much} & \text{Nmr:Sg} & (1225) \\
\text{many} & \, \text{Det.Indef.Cnt:many} & \text{Nmr:Pl} & (1226) \\
\text{some} & \, (\text{Det.Indef.Mass:some} & \text{Nmr:Sg},) \\
& & \text{Det.Indef.Cnt:some} & \text{Nmr:Pl}) & (1227) \\
\text{any} & \, (\text{Det.Indef.Mass:any} & \text{Nmr:Sg},) \\
& & \text{Det.Indef.Cnt:any} & \text{Nmr:Pl}) & (1228)
\end{align*}
\]
The count (Cnt) or mass (Mass) suffix on the indefinite determiners constrains their co-occurrence in phrase structure with similarly differentiated nouns (cf.4.5).

It should be noted that the entries for 'some' (1227) and 'any' (1228) are designed to handle these elements in their unstressed incidences (Quirk and Greenbaum 1975:62, 69), which are probably the more frequently occurring types. Some examples of unstressed 'some' and 'any' are in 'I don't want any pencils', and 'I want some fruit'.

Number distinguishes when one of these elements 'some' or 'any' co-occurs with a mass noun or a count noun (cf.4.5). A count noun such as 'pencil' must be in its plural form to co-occur with unstressed 'some' or 'any' (cf.2132); thus, plural number (Nmr:Pl) and count (Cnt) are specified in their entries. A mass noun such as 'fruit' is always characterized as singular (except under special circumstances as in 'He bought three beers' meaning 'He bought three bottles/cans/glasses of beer', but we ignore these here), so singular number (Nmr:Sg) is specified in the entries for 'some' and 'any' when they co-occur with mass nouns (cf.2133).

Stressed 'some' and 'any' also introduce mass or count nouns, but the count noun can be in its singular form as in 'Any bottle is better than none', or 'That's some pen you have there' (Quirk and Greenbaum 1975:62,109 fn). As indicated above, however, these stressed uses are not presently accommodated in the grammar.
Thus far, we have been looking at determiners in their specific reference capacity. Specific reference is to one or more particular items or individuals. There is, however, another system, that of generic reference. According to Quirk and Greenbaum, a reference is generic when we are indicating a class of things. '...Generic reference is used to denote what is normal or typical for members of a class.' (Quirk and Greenbaum 1975:68) Their examples are:

The tiger is a dangerous animal.
A tiger is a dangerous animal.
Tigers are dangerous animals. (1975:68)

Quirk and Greenbaum state that '...the distinctions of number and definiteness are neutralized since they are no longer relevant for the generic concept.' (1975:68) For this reason, 'a' and 'the' have a separate classification, that of generic (Gen). Entries (1200) and (1220), shown only in part above, actually have the following forms:

\texttt{t h e \ (Det.Def, Det.Gen):the \ Nmr:(Sg,Pl)} \quad (1200)
\texttt{a \ (Det.Indef.Cnt, Det.Gen):a \ Nmr:Sg} \quad (1220)

Number is neutralized when a reference is generic in the sense that singular and plural terms refer to the same thing, a class. So, 'a tiger', 'the tiger', and 'tigers' all denote the class tiger. Number is still necessary, however, since it restricts the co-occurrence of these elements (cf.4.6).

3.3 Pronouns

Our grammar includes personal pronouns and relative and interrogative pronouns. The personal pronouns are repre-
sented as the numerals '1', '2', or '3', which correspond to
their person, and their number (plurality) is also shown.

The entries for the personal pronouns are:

\begin{align*}
\text{I} & \ \text{Pro:1} \ \text{Nmr:Sg} \quad \text{(1100)} \\
\text{w e} & \ \text{Pro:1} \ \text{Nmr:Pl} \quad \text{(1101)} \\
\text{y o u} & \ \text{Pro:2} \ \text{Nmr:(Sg,Pl)} \ (\emptyset,\text{m}) \quad \text{(1102)} \\
\text{h e} & \ \text{Pro:3-male} \ \text{Nmr:Sg} \quad \text{(1103)} \\
\text{s h e} & \ \text{Pro:3-female} \ \text{Nmr:Sg} \quad \text{(1104)} \\
\text{i t} & \ \text{Pro:3-neuter} \ \text{Nmr:Sg} \ (\emptyset,\text{m}) \quad \text{(1105)} \\
\text{t h e y} & \ \text{Pro:3} \ \text{Nmr:Pl} \quad \text{(1106)} \\
\text{m e} & \ \text{Pro:1} \ \text{Nmr:Sg} \ \text{m} \quad \text{(1107)} \\
\text{u s} & \ \text{Pro:1} \ \text{Nmr:Pl} \ \text{m} \quad \text{(1108)} \\
\text{h i m} & \ \text{Pro:3-male} \ \text{Nmr:Sg} \ \text{m} \quad \text{(1110)} \\
\text{h e r} & \ \text{Pro:3-female} \ \text{Nmr:Sg} \ \text{m} \quad \text{(1111)} \\
\text{t h e m} & \ \text{Pro:3} \ \text{Nmr:Pl} \ \text{m} \quad \text{(1113)} \\
\end{align*}

The 'm' in several of the entries is the objective case
marker.

The relative and interrogative pronouns are:

\begin{align*}
\text{w h o} & .. \ \text{(Relv, Intr)} .. \ \text{Pro:who} \quad \text{(1150)} \\
\text{w h o m} & .. \ \text{(Relv, Intr)} .. \ \text{Pro:who m} \quad \text{(1151)} \\
\text{w h i c h} & .. \ \text{(Relv, Intr)} .. \ \text{Pro:which} \quad \text{(1152)} \\
\text{t h a t} & .. \ \text{Relv} .. \ \text{Pro:that} \quad \text{(1153)} \\
\text{w h a t} & .. \ \text{Intr} .. \ \text{Pro:what} \quad \text{(1154)} \\
\end{align*}

The graphemic string is followed by an optional, un-
specified string (..). The graphemic string is the realiza-
tion of two elements: one is the relative (Relv) or inter-
rogative (Intr) element, and the other is that specified to
the right of the colon. The optional, unspecified string
(..) which intervenes between the two elements signifies
that those elements may be discontinuous (cf. 2.2, 2.4, 7.2,
and 7.3). In the sentence, 'I wanted to wear the pointe
shoes which I had just bought', the relative pronoun
'which', though occurring at the beginning of the relative
clause, serves as the object of the verb 'buy'. In unmarked
word order, the elements 'Relv', which marks the beginning
of the relative clause, and 'which' are discontinuous, with intervening 'I had just bought'. Entry (1152), instantiated to the above relative clause, is as follows:

```
which (I had just bought) \ Relv (I had just bought) Pro:which
```

The portions in parentheses correspond to the two dots in the rule.

When the relative or interrogative pronoun is acting as subject of the clause, then the two dots are realized as zero. For example, in 'Who ate the pie?', 'Intr', which marks the beginning of the interrogative clause, is immediately followed by Pro:who, which is the subject. Rule (1150) instantiated to this example is:

```
who \ Intr Pro:who
```

The entry for the relative pronoun 'that' has the same entry number as that for the determiner 'that' (cf.3.2). These two 'thats' are entered in the Morphicon only once, signifying that they are homographs (homonyms). The different tactic codes for the two (shown on the content side of the rules) distinguish them as different morphemes.

The plurality of relative and interrogative pronouns is not specified at this level. Their number is based on that of the noun they co-occur with in constructions. If there is no such noun, which is a distinct possibility for interrogative pronouns, they are marked as singular (cf.4.7,4.8).

### 3.4 Numerals

The numerals are categorized as either cardinal or
ordinal according to traditional criteria. Entries for a few of the numerals in the Morphicon are:

\[
\begin{align*}
on \ e & \ \text{Num.Card:one Nmr:Sg} & (1600) \\
1 & \ \text{Num.Card:one Nmr:Sg} & (1601) \\
\text{fi}r\text{s}t & \ \text{Num.Ord:first} & (1650) \\
\text{sec}\text{on}d & \ \text{Num.Ord:second} & (1651)
\end{align*}
\]

The cardinal numerals accommodated in the grammar may be symbols, '1, 2, 3,' etc., or words, 'one, two, three,' etc.

Ordinal numerals do not have any innate plurality as cardinals do. Consequently, no number is associated with them as it is for cardinals.

3.5 Adjectives

Adjectives are divided into sub-categories. The distributional criteria for the division is that a particular sequence of these sub-types produces the unmarked order. The semantic explanation often given for this ordering is that these sub-types reflect some 'perceived relative closeness of the modifier (adjective) to the head' (the noun modified by the adjective) (Fries 1986:132).

The five sub-categories of adjectives are: subjectively measured, objectively measured, age, color, and classifier. These categories come primarily from Quirk and Greenbaum (1975:125), although they do not correspond exactly to their system. One exception is that of classifier which is a term borrowed from Halliday (1985:164) to cover Quirk and Greenbaum's two categories of 'denominal adjectives denoting (either) material, eg: a woolen scarf, ... and ... resem-
blance to a material, eg: ...silken hair, ... (or) ...
provenance or style, eg: a British ship, a Parisian dress'
(1975:125).

Halliday states that 'classifiers do not accept degrees
of comparison ...' and that 'the range of semantic relations
functioning as classifier is very broad: it includes
material, scale and scope, purpose and function, status and
rank, origin, mode of operation - more or less any feature
that may serve to classify a set of things into a system of
smaller sets.' (1985:164) Some entries for classifier
adjectives are:

\begin{verbatim}
metallic \ Adj.Clas:metallic (1780)
woolen \ Adj.Clas:woolen (1781)
domestic \ Adj.Clas:domestic (1782)
electric \ Adj.Clas:electric (1783)
\end{verbatim}

Some of the 'general adjectives susceptible to
subjective measure ...' (Quirk and Greenbaum 1975:125)
included in the Morphicon are:

\begin{verbatim}
sick \ Adj.Subj:sick (1700)
happy \ Adj.Subj:happy (1702)
funny \ Adj.Subj:funny (1704)
good \ Adj.Subj:good (1705)
superior \ Adj.Subj:superior (1706)
\end{verbatim}

Some adjectives denoting shape or size, which can be
objectively measured (Quirk and Greenbaum 1975:125) are:

\begin{verbatim}
round \ Adj.Obj:round (1720)
little \ Adj.Obj:little (1721)
\end{verbatim}

Just as some nouns are classified as both count and
mass (e.g. 'time' cf.3.1), so too are some adjectives sub-
categorized as both subjectively and objectively measured.
Take, for example, 'little' in 'She's just a little girl'
and 'It's just a little disaster'. The semantic explanation is that its use in the former statement denotes a physical measurement ("the relation between a particular item and the norm for the class to which that item has been assigned" [Fries 1986:129, fn.6]), while its use in the latter projects the attitude of the speaker (Fries 1986:129). 'Little' is thus categorized as both objectively and subjectively measured.

\[
\text{l i t t l e} \ \text{Adj.(Subj,Obj):little} \quad (1721)
\]

Some adjectives of age and color are:

\[
\text{o l d} \ \text{Adj.Age:old} \quad (1740) \\
\text{r e d} \ \text{Adj.Col:red} \quad (1760)
\]

3.6 Adverbs

According to Gleason, 'the "adverb" of school grammar covers a remarkably diverse group of words.' (1965:120) He recommends excising several categories from this very heterogeneous group. One new category he suggests is that of 'intensifier'. Included in this group are elements such as 'very' or 'rather' (1965:130-1). Some of these elements may have other labels as well.

\[
\text{v e r y} \ \text{Intns:very} \quad (1940) \\
\text{r e a l l y} \ \text{Intns:really} \quad (1941) \\
\text{q u i t e} \ \text{Intns:quite} \quad (1942) \\
\text{p r e t t y} \ \text{Intns:pretty} \quad (1943)
\]

Quirk and Greenbaum list additional intensifiers - 'so', 'unusually', 'unbelievably', and the more informal 'that' as in 'He's not that old!' (1975:127).

Like intensifiers, 'not' is sorted out from other
adverbs. The uniqueness of its occurrences earns it a
category all its own. Since this is the only element in the
category, no tactic label is used.

\[
\text{not} \quad \text{(1950)}
\]

'Not' has two realizations, the one just shown and the
one that follows.

\[
(Aux:X, \text{Mod}:Y \quad [\text{Tns}:X1], \text{V}:Be:be) \quad \text{-n ' t \ (Aux:X, \text{Mod}:Y}
\quad [\text{Tns}:X1], \text{V}:Be:be) \quad \text{not} \quad \text{(1987)}
\]

This contracted form appears as a suffix on
auxiliaries, modals, and the verb 'be'.

The remainder of the adverbs, those of manner,
frequency, location, time, degree, and probability, are
subsumed by the general category adverb, with sub-
categories. An entry of each sub-type follows.

\[
\begin{align*}
\text{here} & \quad \text{Adv.Loc:here} \quad \text{(1800)} \\
\text{almost} & \quad \text{Adv.Degr:almost} \quad \text{(1821)} \\
\text{now} & \quad \text{Adv.Temp:now} \quad \text{(1840)} \\
\text{always} & \quad \text{Adv.Freq:always} \quad \text{(1860)} \\
\text{easily} & \quad \text{Adv.Manr:easily} \quad \text{(1893)} \\
\text{certainly} & \quad \text{Adv.Prob:certainly} \quad \text{(1880)}
\end{align*}
\]

Locative adverbs indicate 'where': 'here', 'there',
'outside', 'nearby'. Frequency adverbs tell 'how often'.
Like 'always' are 'sometimes', 'never', 'rarely', and
'often'. Some temporal adverbs that specify 'when' are
'now', 'then', 'today', and 'formerly'. One sub-category of
temporal adverb, 'o'clock' (cf.1844), occurs only in a
construction with the numbers one through twelve (cf.4.20).
Degree adverbs indicate near completeness. Like 'almost'
are 'practically' and 'nearly'. Manner adverbs tell 'how'
something is done: 'easily', 'rapidly', 'cautiously',
'respectfully'. Probability adverbs such as 'certainly', 'probably', and 'possibly' indicate modality, the area intermediate to yes and no (Halliday 1985:335). The meaning of these probability adverbs is explored further in 5.5.

The above sub-categorizations are based on a preliminary analysis of their distribution. The present grammar does not try to give definitive evidence for every one of these divisions in the form of lextactic constructions, although some are specified (cf.4.16,4.18,5.4,5.5).

In addition to the adverbs listed above, there are interrogative adverbs.

```
\how\ \Intr\ \Adv.Manner:how (1920)
\where\ \Intr\ \Adv.Loc:where (1921)
\when\ \Intr\ \Adv.Temp:when (1922)
```

The graphemic string on the left is the realization of two elements, interrogative and the adverb element. This treatment mirrors that given to interrogative and relative pronouns (cf.3.3). The two dots stand for that which may intervene between the two elements (cf.2.2,2.4,7.3).

The interrogative adverb 'when' is illustrated by the following question and its instantiation in rule form (cf. 4.19).

'When are we leaving?'

```
when (are we leaving) \Intr (we are leaving) Adv.Temp:when
```

3.7 Prepositions

Some of the prepositions that are included in the grammar are:
The present grammar does not distinguish sub-types of prepositions since all prepositions covered here have the same lexotactic co-occurrence restrictions, preposition followed by noun phrase (cf.4.9). At a more delicate degree of analysis, however, some lexotactic constructions may exclude certain prepositions which would necessitate a division into sub-types.

3.8 Verbs

The entries for verbs are divided into the main categories of auxiliaries, modals, and verbs. Verbs are further subdivided into transitive, intransitive, copulas, and 'be'.

Since many verbs have the function of more than one type, they are assigned to multiple categories. For example, 'do' is an auxiliary in questions and negatives: 'Do you dance' or 'I don't dance'; a transitive verb in 'I do my homework'; and an intransitive verb in 'He does well'.

The entries for 'do' follow.

\[\text{do} \ (\text{Aux}, \ V.T, \ V.I) : \text{do} \quad (1410)\]
\[\text{does} \ \text{C:3-Sg Tns:Pres} \ (\text{Aux}, \ V.T, \ V.I) : \text{do} \quad (1411)\]
\[\text{did} \ \text{Tns:Past} \ (\text{Aux}, \ V.T, \ V.I) : \text{do} \quad (1412)\]
\[\text{done} \ \text{en} \ (V.T, \ V.I) : \text{do} \quad (1413)\]

As explained for nouns in 3.1, many inflected forms are placed in the Morphicon as single units. The graphemic string "does" is the realization of three elements: the third person singular marker (3-Sg), present tense
(Tns:Pres), and 'do'. Likewise, "d i d" realizes 'past' tense and 'do'. "D o n e" is the graphemic manifestation of the past participle 'en' and 'do'.

Due to the special status of 'be' in the language, it is given its own sub-category of verb (V.Be). All the forms of 'be' are given below.

\[
\begin{align*}
\text{b e} & \quad \text{V.Be:be} & (1400) \\
\text{b e i n g} & \quad (i n g \ V.Be:be, \ N.Cnt:being) & (1401) \\
\text{b e e n} & \quad \text{V.Be:be} & (1402) \\
\text{i s} & \quad \text{C:3-Sg} \ Tns:Pres \ V.Be:be & (1403) \\
\text{a m} & \quad \text{C:1-Sg} \ Tns:Pres \ V.Be:be & (1404) \\
\text{a r e} & \quad \text{Tns:Pres} \ V.Be:be & (1405) \\
\text{w a s} & \quad \text{C:(3-Sg,1-Sg)} \ Tns:Past \ V.Be:be & (1406) \\
\text{w e r e} & \quad \text{Tns:Past} \ V.Be:be & (1407)
\end{align*}
\]

Some examples of transitive verbs (V.T) are 'eat', 'take', 'kick', and 'want'. Their entries are like this one for 'eat'.

\[
\begin{align*}
\text{e a t} & \quad \text{V.T:eat} & (1460)
\end{align*}
\]

Intransitive verbs (V.I) such as 'go', 'live', 'sit', and 'sleep' belong in entries such as this one.

\[
\begin{align*}
\text{g o} & \quad \text{V.I:go} & (1550)
\end{align*}
\]

Copulas (V.C) are also often called linking verbs.

\[
\begin{align*}
\text{b e c o m e} & \quad \text{V.C:become} & (1440)
\end{align*}
\]

Like 'become' are 'seem', 'feel', 'act', 'sound', 'look', and 'taste'. 'Become' behaves somewhat differently than the other copulas. It is the only one (at least in the dialect being presented) that may have a noun phrase as complement, for example, 'He became a doctor', but '?He looked a doctor' (cf.4.16 (2204)). In 4.20, a rule for lexemes that consist of these copulas together with the
preposition 'like', for example, 'look-like', is given. It is that complex copula construction which takes a noun phrase as complement, for example, 'He looked like a doctor'.

-s, or its alternate form -e s, is the realization of the third person singular present tense marker in written English. Two entries accommodate these graphemic strings.

\[ V.X:Y \ -s \ \ C:3-Sg \ Tns:Pres \ V.X:Y \quad (1982) \]

-s, occurring as a suffix on a verb Y of any sub-type X, is the realization of a concord element third person singular and a present tense element.

\[ V.X:Y<-s, \ -z, \ -x, \ -ch, \ -sh, \ -y/i> \ -e s \ \ C:3-Sg \ Tns:Pres \ V.X:Y \quad (1983) \]

This statement says that the suffix -e s, occurring as a suffix on a verb listed in the dictionary as ending in s, z, x, ch, sh, or y, has the same realizes as those given for the suffix -s above.

Entry (1982) provides for verb forms such as "eats" in 'He eats meat', while (1983) allows for "tries" in 'He tries hard'.

Like -s and -e s, there are two alternate realizations, -d and -e d, of the past tense or past participle element.

\[ V.X:Y \ -d \ (Tns:Past, \ en) \ V.X:Y \quad (1984) \]

Entry (1984) states that -d, occurring as a suffix on a verb of any sub-type, is the realization of either past tense (Tns:Past) or the past participle (en). Some example phrases whose verb this rule allows for are: 'He tied them to a bedpost' and 'He was tied behind the barn'.
(1985) presents essentially the same information that (1984) does, but it expresses spelling considerations.

\[ V.X: (Y, Y<-y/i>, Y<-X1/X1X1>) -e \ d \ \ (Tns:Past, \ en) \]
\[ V.X:Y \ IF \ X1 = (b,d,g,p,t,m,n,l,r) \]  

(1985)

This rule has three parts. The left side of the first part is: \( V.X:Y -e \ d \ \), for verbs with the suffix \(-e \ d\), as in "wanted" (cf.1466).

The second part deals with the \( y/i\) spelling alternation: \( V.X:Y<-y/i> -e \ d \ \). A verb that ends in \( y\) in the dictionary appears with \( i\) rather than \( y\) before \(-e \ d\) in its graphemic realization. "Cried", as in 'She cried daily', exemplifies the type of form accommodated by this statement.

The third part of the rule considers consonant doubling: \( V.X:Y<-X1/X1X1> -e \ d \ \ ... \ IF \ X1 = (b,d,g,p,t, m,n,l,r). \) This portion of the entry states that a verb listed in the dictionary as ending in \( b,d,g,p,t,m,n,l,\) or \( r\), in its graphemic realization has double \( b\), double \( d\), double \( g\), etc. before \(-e \ d\). "Hopped" as in 'The bunny hopped along' is provided for by this rule.

The rule that takes care of the suffix in present participle (ing) forms of verbs is:

\[ V.X: (Y, Y<-e/\emptyset>, Y<-X1/X1X1>) -i \ n \ g \ \ ing \ V.X:Y \]
\[ IF \ X1 = (b,d,g,p,t,m,n,l,r) \]  

(1986)

Some of the provisions of this rule resemble those found in entry (1985) just discussed. The part that differs concerns the spelling convention of "dropping" the final \(-e\) before "adding" \(-ing\). This feature is expressed by the portion of the rule: \( V.X:Y<-e/\emptyset> -i \ n \ g \ \). A verb that
ends in e in the dictionary appears without the e in its
graphemic representation when it co-occurs with the suffix
-ing. "Hoping", as in 'We are hoping to go', illustrates
this rule.

Modals constitute the third major category of verbs.
They are treated quite differently from other verb types,
however, as is explained more fully in 5.5. A few of the
English modals are:

\begin{verbatim}
  will \ Mod:will (\emptyset, Tns:Fut)  \hspace{1cm} (1420)
  can \ Mod:can \hspace{1cm} (1422)
  could \ (Mod:can Tns:Past, Mod:could) \hspace{1cm} (1424)
\end{verbatim}

The graphemic string "will" is the realization of
either just the modal 'will' or the modal 'will' and
'future' tense. This choice is expressed in the separation
by a comma of the elements zero (\emptyset) and future tense
(Tns:Fut), and by their enclosure in parentheses. To
appreciate the different categorizations of 'will', we
appeal to a semantic explanation and offer as an example:
'I will wash the windows'. In one context 'will' means
something akin to 'agree to' or 'volunteer'; in another it
signifies future time for the verb it accompanies (in this
case 'wash').
NOTES

CHAPTER 3

1. From personal communication with Lamb. Cf. also Hockett (1987:114), who proposes that the strategies of the hearer also serve as those of the speaker since 'we hear our own speech as well as that of others, so that what we transmit is a proper subset of what we receive. Furthermore, we understand our own speech by just the mechanisms by which we understand the speech of others; we can know what we are saying only because we hear it.'

2. Some other rules have been worked out for the lextactic constraints of adverbs, but not included in the present grammar. Some of these rules are:

   Adv.Degr:X Adv.Freq:(always,never) \ AdvPhr:(Adv.Degr:X Adv.Freq:(always,never)) e.g. 'almost always'
   Intns:X Adv.Freq:(often,frequently,seldom) \ AdvPhr:
     (Intns:X Adv.Freq:(often,frequently,seldom))
   e.g. 'very often'
CHAPTER 4
THE LEXICON

4.0 Entries of the Lexicon

The Lexicon is intermediate to the morphemic and lexemic levels of linguistic structure (cf. 2.1). The entries in the Lexicon, as distinct from those in the Morphicon, are of two main types: those for lexemic idioms, and those which provide for general phrase structure. Formally, the difference is that the entries of the latter type have variables.

4.1 Idioms

Idioms are made up out of the elements of language of the type described in Chapter 3. Some examples of idioms are 'cat-nap', 'old-money', 'head-over-heels', 'look-at', 'put-up-with', and so forth.

Idioms have specific entries in the Lexicon, just as simple words have in the Morphicon. The literal interpretations of these idioms are taken care of by general phrase structure rules. Some entries for idioms included in the Lexicon are:

\[
\text{N.Cnt:cat N.Cnt:nap} \ \text{\textbackslash N.Cnt:cat-nap} \quad (2030)
\]
\[
\text{Adj.Age:old N.Mass:money} \ \text{\textbackslash N.Mass:old-money} \quad (2031)
\]
\[
\text{V.I:sit Prep:in} \ \text{\textbackslash N.Cnt:sit-in} \quad (2032)
\]

An idiom may have as its realization a combination of elements, but it is a single element at the lexemic level,
hence the use of the hyphen(s) in the lexeme symbol (cf. 2.2). 'Cat' and 'nap' are two separate morphemes, but 'cat-nap' is one lexeme.

V.T:take .. Prep:on \ V.T:take-on .. (2002)

The above entry states that 'take' and 'on', following 'take' though not necessarily immediately following it (indicated by the two dots (cf.2.2,2.4)), can constitute a realization of the idiom 'take-on', meaning 'assume' or 'contend with'. Some examples of 'take-on' in use are: 'He took on the team' and 'He took the team on'. In both instances, 'take-on' is a single lexeme. In the latter example, the object noun phrase 'the team' intervenes between the morphemes in the realization of 'take-on'. The two dots on the left side of this rule allow for this intervention. In the former example, 'take' and 'on' are immediate constituents; the two dots are therefore realized as zero.

A further entry of this type follows.


Unlike the idiom 'take-on', discussed above, 'look-in-on' can not be realized as discontinuous morphemes; that is, nothing may intervene between its parts. Consequently, the two dots do not appear in this rule as they do in (2002).

Although we have mentioned only a few, combinations of elements of all parts of speech can be the realization of a lexemic idiom in English.
4.2 **Number (Plurality)**

Nouns like 'men' and 'women', inflected forms that are entered as single units in the Morphicon, have plural number (Nmr:Pl). In addition, those nouns which are inflected for plural in the usual manner, with an -(e)s suffix, are also thus marked for number (cf.3.1). Any noun that is not marked as plural in one of these ways is by default marked as singular, by virtue of the following entry in the Lexicon.

\[
\text{N.X.Y Nmr:(Pl,0) } \backslash \text{ N.X.Y Nmr:(Pl,Sg)} \quad (2100)
\]

The order of the elements in parentheses on one side of the rule corresponds to that on the other side. Plural is, therefore, the realization of plural, and zero is the realization of singular.

Number restricts nouns in their co-occurrences with other elements. The specific restrictions are discussed in 4.5, 4.6, and 4.13 below.

4.3 **Adjective phrases**

An adjective phrase (AP) consists of an adjective together with any modifiers it may have.

\[
\]

The optional modifying elements in an adjective phrase are adverb (Adv) or intensifier (Intns). The tactic codes of all the elements in the adjective phrase are present on the right side of this entry since they are necessary for higher level entries. Their usefulness is apparent in 4.4,
5.2, 5.4, and in processing of information (cf.7.1).

Some examples of adjective phrases are: 'sick' (adjective only), 'very sick' (intensifier + adjective), 'almost sick' (degree + adjective), 'always sick' (frequency + adjective), 'now sick' (temporal + adjective), 'possibly sick' (probability + adjective).

Manner and locative adverbs (cf.3.6) do not usually occur as modifiers of adjectives although a context for their occurrence could always be conjured. We do not, therefore, specifically exclude them in this rule; for example, ?'The here sick, there well man forgot where he was and acted energetic' (locative), and ?'The responsibly sick man took his pills' (manner).

The degree of delicacy to which the modification in this entry is taken is rather less than that for some other phenomena in this grammar. The reader will find that this is the case for all of the constructions which include adverbs (cf.3.6, Chapter 3:fn.2).

Adjective phrases may be prenominal or predicative. (Postnominal adjectives, as in 'secretary general', are not addressed in the present grammar. Neither are postnominals which are reduced clauses or phrases, as in 'She likes things expensive'.) An example with a prenominal adjective phrase is 'the completely wrong answer'. An example of predicative use is 'the answer is completely wrong'. The adverb in this pair of examples is modifying the adjective, as provided for in (2120). In a pair of examples such as 'the frequently
sick girl' and 'the girl is frequently sick', however, the
adverb in the former is modifying the adjective, while in
the latter, it is probably construed to be modifying the
verb. Further investigation into the lexotactic co-occurrence restrictions on adverbs is needed to specify exactly
which adverbs, in a be/copula-predicate adjective construc-
tion, belong to the verb and which belong to the adjective.
2
For now, we leave rule (2120) as is and bind an adverb im-
mediately preceding an adjective to that adjective in the
adjective phrase construction. The occurrence of any other
adverbs is then covered in the rule for verb phrases and
predicates (cf. 4.18, 4.19).

4.4 Nominals

A nominal has a head and optional modifiers. There are
two entries in the Lexicon for nominals. The first entry
shown has a pronoun as the head of a nominal.

Pro:X Nmr:Y \ Nom:(Head:(Pro:X Nmr:Y))  (2126)

This rule instantiated to the pronoun 'I' (cf. 3.3) is:

Pro:1 Nmr:Sg \ Nom:(Head:(Pro:1 Nmr:Sg))

In the second entry, the right side of the rule shows
that the head of the nominal consists of the noun element
and its plurality (cf. the end of the last line in the
rule). The modifier (Modif) includes any or all of the
adjective phrases.
Some examples of nominals accommodated by this second rule are 'happy Harry', 'water', and 'cute little kitties'.

The adjective phrases occur in a particular order with respect to one another, indicated by their sequence in the entry (cf. 3.5 and Quirk and Greenbaum 1975:125). Subjective adjectives precede objective adjectives, which precede age adjectives, which, in turn, precede color adjectives, which precede classifier adjectives. This is the unmarked order for adjectives in English. An example of a nominal with an adjective of each sub-type is '(the) quaint little old red wooden barn'. The right side of the above rule, instantiated to this example is:


Because each sub-type of adjective can occur more than once, the sign for repetition (the asterisk) is placed after each adjective phrase in the entry. An example of a repeated sub-type is found in an utterance such as 'The clerk gave the man a peculiar, superior look', wherein both 'peculiar' and 'superior' are elements of the sub-type subjective.

Some questionable combinations of elements are allowed by this rule; antonyms such as 'hot cold' or 'old young',

```plaintext
[..Adj.Age: X1]* [..Adj.Col: Y1]*  
[..Adj.Clas: X2]*)]  Head: (N.Y2:X3  Nmr: Y3))
```
and other conceptual inconsistencies such as 'square dome'. A bit of imagination might allow us to find instances, however, in which some of these apparent contradictions are compatible. In any case, these questionable combinations are resolved at a level higher than that of the Lexicon, so they do not affect this entry (cf. 5.2).

4.5 Noun phrases with specific reference

There are seven entries that describe noun phrases with specific reference. Specific denotes a reference to one or more members of a class. The first entry has a proper noun as head.

\[
[\text{Num.Card}: X \text{ Nmr}: Y] \quad \text{Nom:} (\ldots \text{Head: (N.Prop: X1 Nmr: Y) \ Np: ([X] \ldots \text{Head: (Y1 Nmr: Y)})})
\] (2130)

Phrases such as 'February', 'cold February', or 'one very snowy February' are accommodated by this rule. It also covers the noun phrase in the sentence given earlier in 3.1: 'There are three Dougs in my ballet class'.

The plurality of the cardinal numeral agrees with that of the proper noun. This agreement is indicated on the left side of the rule by the appearance of the variable Y as the number of both numeral and noun. On the right side of the entry, plurality is expressed only once, on the head noun.

The cardinal numeral is an optional element, as indicated by its enclosure in the square brackets [ ].

The two dots in this and the following rules for noun phrases stand for the modifier (cf. rule for nominal in 4.4). The tactic labels for the numeral (Num.Card) and noun
type (N.Prop) are not stated when those elements are part of a noun phrase. (Cf. their absence on the right side of the rule.) There are no other constructions whose co-occurrence restrictions depend on this information. The reader should remain aware that this is true for all of the noun phrase constructions although we do not repeat it below in the discussion of each one.

The rule instantiated to 'one very snowy February' is:

\[\text{Num.Card:one Nmr:Sg Nom: (Modif: (Intns:very Adj.Clas:snowy)) Head: (N.Prop:February Nmr:Sg)) \backslash NP: (one Modif: (Intns:very Adj.Clas:snowy) Head: (February Nmr:Sg))}\]

The next entry for noun phrases is concerned with those elements which co-occur with the definite determiners.

\[\text{Det.Def:X Nmr:Y ~(}[\text{Num.Card:X1 Nmr:Y}] \text{ [Num.Ord:Y1]}\text{ Nom: (..Head: (N.X2:Y2 Nmr:Y)) \backslash NP: (X ~(}[\text{X1} [Y1]] .. Head: (Y2 Nmr:Y))\text{ (2131)}}\]

Number (plurality) agreement in this entry is similar to that discussed for the previous rule, but here there are three constituents in accord with one another: the determiner, the cardinal numeral, and the head noun (cf. the variable Y).

The two non-optional constituents of this rule are the definite determiner and the head noun. This entry shows that definite determiners co-occur with any of the noun sub-types, count, mass, proper, or even hour (cf.4.20), since the noun sub-type specified is a variable, X2. Noun phrases such as these are thus accommodated by this entry: 'those lunatics' (plural count noun), 'that funny money' (mass
noun), 'the Houston' (as in 'The Houston that I know is a very clean city.') (proper noun), and 'the ten o'clock' (as in 'I took the ten o'clock to town') (hour noun).

Both ordinal and cardinal numerals may appear in a definite noun phrase. That they may occur in any order with respect to one another is indicated by their inclusion in parentheses preceded by a tilde (~) (cf.2.2). Some examples of definite noun phrases with numerals are: 'the one wonderful story', 'those two first cars', and 'these first two evenings'. The representation of the last example in rule form is:

\[
\begin{align*}
\text{Det.Df:} & \text{this Nmr:Pl Num.Ord:} \text{first} \\
\text{Num.Card:two Nmr:Pl Nom:(Head:(N.Cnt:evening Nmr:Pl))} \quad \text{NP:} & \text{(this first two Head:(evening Nmr:Pl))}
\end{align*}
\]

Indefinite determiners, sub-categorized as count or mass, enter into noun phrases with head nouns that are similarly differentiated. The entry for sub-type count is:

\[
\begin{align*}
\text{Nom:} & \text{(..Head:(N.Cnt:Y1 Nmr:Y))} \\
\text{NP:} & \text{(X [X1]..Head:(Y1 Nmr:Y))} \quad (2132)
\end{align*}
\]

As above, there is agreement among constituents, in this case the determiner and noun. They are in accord in their number (cf. the variable Y) and their sub-category (cf. the suffix Cnt on their tactic labels).

Some examples of phrases provided for by this entry are: 'each mitten', 'many busy bees', 'a second pie', 'some little kitties', and 'every third morning'.

This entry admits only ordinal numerals. Cardinal numerals can also co-occur with the indefinite count
determiners 'every' and 'each' as in 'every two months' or 'each five lessons'. They can not, however, co-occur with many of the other indefinite determiners (cf. *a one person, *many five men) (Quirk and Greenbaum 1975:65,108). When 'each' and 'every' do occur in combination with plural expressions such as 'two months', that expression is functioning like a single unit; that is, 'two months' is a unit of time like 'month' (cf.'every month') or 'day' (cf.'every day'). Because it is a unit, it is functionally singular, the plurality also associated with 'each' and 'every' (cf.3.2 (1221,1222)). The entry to cover the co-occurrence of cardinal numerals with 'each' and 'every' is not included in the present grammar.

The entry for noun phrases with indefinite determiners and mass nouns is:

```
Det.Indef.Mass:X Nmr:Sg Nom:(..Head:(N.Mass:Y Nmr:Sg))
 | NP:(X..Head:(Y Nmr:Sg))
```

(2133)

Phrases provided for by this entry are: 'much cold water', 'some sticky candy', and 'any money'.

The suffix Mass on the determiner and on the head noun, as well as singular (Sg) number on both, show the concord that is necessary between these constituents in order to group them under the label noun phrase.

The next entry also provides for an indefinite noun phrase, but there is no determiner introducing the noun.

```
 | NP:([X]..Head:(Y Nmr:Pl))
```

(2136)

This entry is for plural count nouns with or without a
cardinal numeral. Representative of the type of noun phrase accounted for by this rule are: 'five hungry cats', 'hungry cats', 'five cats', or just 'cats' in a sentence such as 'I see five hungry cats'.

A mass noun or hour noun also occurs without an article, but denotes indefiniteness nonetheless.

\[
\begin{align*}
\text{Nom:} & \cdot \text{Head:} (N. (\text{Mass, Hour}): X \text{ Nmr: Sg}) \setminus \\
\text{NP:} & \cdot \text{Head:} (X \text{ Nmr: Sg}) \\
\end{align*}
\]

(2137)

'Good furniture' and 'nine o'clock', as in 'They have good furniture' and 'It's nine o'clock' are illustrative of the noun phrases taken care of by this entry.

The following entry shows a noun phrase with a pronoun as its head.

\[
\begin{align*}
\text{Nom:} & \cdot \text{Head:} (\text{Pro}: X \text{ Nmr: Y}) \setminus \\
\text{NP:} & \cdot \text{Head:} (X \text{ Nmr: Y}) \\
\end{align*}
\]

(2134)

Pronouns may occur alone as noun phrases. This entry is for personal pronouns (cf. 4.7 for other types of pronouns). Personal pronouns are always definite, even more so than noun phrases introduced by definite determiners. The semantic explanation for this definiteness has to do with the fact that the hearer must have the referent of the pronoun still in mind (Copeland and Davis 1981:128).

4.6 Noun phrases with generic reference

Generic reference is to a class rather than to a particular item or individual. To indicate this type of reference, an element 'generic' (Gen) is present in the noun phrase.

This first entry presented takes care of generic
references that are introduced by 'the' or 'a'.

\[
\text{Det.Gen:X Nmr:Sg Nom:} (\text{..Head:} (\text{N.Cnt:Y Nmr:Sg})) \\
\text{NP:} (\text{Gen..Head:} (\text{Y Nmr:Sg})) \quad (2140)
\]

For examples of this type of noun phrase, we refer back to those given earlier from Quirk and Greenbaum (1975:68) (cf. 3.2):

The tiger is a dangerous animal.  
A tiger is a dangerous animal.

The reader will notice that this entry provides only for the occurrence of a singular count noun after a generic determiner ('a' or 'the'). Mass nouns and plural count nouns do not co-occur with a determiner when the reference is generic.

The variable X, which stands for 'a' or 'the', is not shown on the right side of the rule. These elements, higher up, signify indefiniteness and definiteness, respectively. This is in keeping with Quirk and Greenbaum's statement, discussed in 3.2, that definiteness is neutralized for a generic concept.

Two other entries in the Lexicon for generic reference are:

\[
\text{Nom:} (\text{..Head:} (\text{N.Cnt:X Nmr:Pl})) \\
\text{NP:} (\text{Gen..Head:} (\text{X Nmr:Pl})) \quad (2141)
\]

\[
\text{Nom:} (\text{..Head:} (\text{N.Mass:X Nmr:Sg})) \\
\text{NP:} (\text{Gen..Head:} (\text{X Nmr:Sg})) \quad (2142)
\]

'Tigers are dangerous animals' (Quirk and Greenbaum 1975:68) illustrates the noun phrase type in (2141), and 'water is wet' exemplifies that in (2142).
4.7 Other noun phrases

There are two other noun phrase entries, both with pronouns as head. The pronoun types are relative and interrogative.

An interrogative pronoun, as described in 3.3, is the realization of two elements: 'interrogative' (Intr) and an element with the tactic classification pronoun (Pro). This pronoun element (X in the rule) participates in a noun phrase as the head.

\[\text{Intr .. Pro:}X \ \text{Intr .. NP:}(\text{Head:}(X \text{ Nmr:Sg})) \] (2155)

An interrogative pronoun is usually singular in number. In one particular construction with a copula and predicate nominative, however, it agrees in number with its antecedent noun (cf.4.16).

'Intr' must be specified in this rule to differentiate the treatment of interrogative pronouns from relative pronouns with regard to their number.

A relative pronoun, like the interrogative pronoun, has two elements: 'relative' (Relv) and an element with the tactic label pronoun (Pro). The pronoun element (X1 in the rule) occurs as a noun phrase, and it carries the plurality associated with its antecedent noun (Y in the rule).

\[\text{NP:} ((\text{..Head:(X Nmr:Y)}) \ [m] \text{ Relv .. Pro:}X1 \ \text{NP:} ((\text{..Head:(X Nmr:Y)}) \ [m] \text{ Relv .. NP:}(\text{Head:}(X1 \text{ Nmr:Y}))) \ [m] \] (2150)

The optional 'm' is the objective case marker (cf.3.3, 4.9,4.16).
4.8 Noun attributes

Noun attributes are all relative clauses, participial phrases, and prepositional phrases that further restrict the referent of a noun. Since the internal composition of these types of attributes is discussed fully in other sections (cf. 4.9, 4.14, 4.17, and 5.2), the reader may simply take these constructions at face value for this discussion and obtain the particulars about them from the subsequent text.

Many prepositional phrases (PP) serve to modify nouns in much the same way that adjectives do. While adjectives are usually restrictive, however, prepositional phrases may be either restrictive or non-restrictive, as these examples show: 'Houston, in a financial crunch, forges on' (non-restrictive) or 'The woman in the park screamed' (restrictive). In the latter example, 'the woman' is a noun phrase and 'in the park' is a prepositional phrase that further restricts the referent designated by 'woman'. The following entry shows, at the higher of the two levels of analysis, the prepositional phrase as part of the noun phrase.

\[ \text{NP:X [m] PP:Y \ \text{NP:} (X \text{ Nattr. PP:Y}) [m] } \] (2501)

The parentheses on the content side shows that NP, which at the lower level is the tactic code for X, is the tactic code for both X and Y.

Participial phrases, like prepositional phrases, help to specify the referent of the noun. For example, in 'the woman wearing the red sweater', 'wearing the red sweater' tells us which woman.
The entry which provides for a participial phrase (PartP or .PtP) as part of a noun phrase is much like the one for prepositional phrases above.

\[ NP:X \ [m] \ PartP:Y \ \ \ NP: (X \ Nattr.PtP:Y) \ [m] \ \ \ (2500) \]

The rule instantiated to the above example is as follows. The internal composition of the participial phrase is not specified in this instantiation:

\[ NP: (\text{the Head:} (\text{woman Nmr:Sg})) \ PartP: (\text{wearing the red sweater}) \ \ \ \ \ \ NP: (\text{the Head:} (\text{woman Nmr:Sg}) \ Nattr.PtP: (\text{wearing the red sweater})) \]

A relative clause (RelvCl or .RC) following a noun phrase is a noun attribute, just as participial and prepositional phrases are.

\[ NP:X \ [m] \ RelvCl:Y \ \ \ NP: (X \ Nattr.RC:Y) \ [m] \ \ \ (2600) \]

Since clauses are not the domain of the Lexicon but of the Semicon, a reference to a clause (cf.RelvCl) at this level denotes a "rankshift": the transfer of a given unit to a lower ...' level (Halliday 1961:251).

Both restrictive and non-restrictive relative clauses are interpreted as noun attributes in the grammar; for example, 'the woman who has on a wig ...' or 'Mrs. Reagan, who has on a wig, ...'. The former example, represented in rule form (but not showing the internal structure of the relative clause) is: \[ NP: (\text{the Head:} (\text{woman Nmr:Sg})) \ RelvCl: (\text{who has on a wig}) \ \ \ NP: (\text{the Head:} (\text{woman Nmr:Sg}) \ Nattr.RC: (who has on a wig)) \]
4.9 Prepositional phrases

A prepositional phrase consists of a preposition followed by a noun phrase. In the prepositional phrase construction, if the head of the noun phrase is a pronoun, then it is marked with the objective case element 'm' required of pronouns in English (cf. the 'm' and IF statement choice in the rule below). Noun phrases whose heads are nouns are unmarked for case (cf. the Ø choice).

\[\text{Prep:} X \ 
\text{NP:}(. . . \text{Head:} (Y \ Nmr: X1) . . )\]
\[\left( \emptyset, m \ \text{IF} \ Y = (1, 2, 3, 3-Y1, \text{who}) \right)\]
\[\sim \text{PP:} (X \ \text{NP:}(. . . \text{Head:} (Y \ Nmr: X1) . . ) \quad (2180)\]

Some example prepositional phrases are: 'to me', 'in the morning', 'on the car', 'at seven o'clock', and 'by the river'. The rule instantiated to 'to me' (cf. 1107) is:

\[\text{Prep:to NP:} (\text{Head:} (1 \ Nmr: Sg)) \ m \ \text{\sim PP:} (\text{to NP:} (1 \ Nmr: Sg))\]

'm' is present on the left side of the rule because Y is instantiated to '1' (cf. the IF statement).

4.10 Tense

As discussed in Chapter 3, various inflected forms of the so-called irregular verbs such as 'be' and 'eat' are entered in the Morphicon with their appropriate tense marker (cf. 3.8). In addition, the so-called regular verbs, which are inflected for tense in the usual fashion with an -(e)d suffix for past or an -(e)s suffix for present third person singular, carry the proper tense marker made possible by entries (1982-1985) in the Morphicon (cf. 3.8). Any verb not already marked in one of these ways is by default marked as present tense by these following two rules in the Lexicon.
Tns: (X, ∅) (Aux, V.X1): Y1 \ Tns: (X, Pres) (Aux, V.X1): Y1

(2110)

Tense (Tns) X is realized as tense X while present tense (Pres) is realized as zero (∅), the absence of an overt marker on the verb. The verbs in the following sentences would fall under the jurisdiction of rule (2110): 'They come every Wednesday' and 'They came every Wednesday'. In the first sentence, no overt tense marker is present, so this verb is in its present tense form (V.I:come \ Tns:Pres V.I:come). In sentence two, we know from an entry in the Morphicon that 'came' is past tense. That interpretation is maintained by this rule (Tns:Past V.I:come \ Tns:Past V.I:come) (cf. Chapter 7: fn. 1).

Mod: X Tns: (Y, ∅) [not] \ Mod: X [not] Tns: (Y, Pres)

(2111)

In (2111), tense Y is realized as tense Y while present tense is realized as zero. The modal (Mod) precedes tense (cf. 3.8) and, on the right side of the rule, 'not' precedes tense since we want all polarity elements (not) and all modality elements (modals) to be at the beginning of the phrase for the purpose of higher level entries (cf. 5.6 & 5.7).

Rule (2111) instantiated to 'couldn't', as in 'Couldn't he see?', is: Mod: could not \ Mod: could not Tns: Pres

Halliday says that 'future', not in combination with some other tense, is realized as 'will' (cf. 3.8 (1420)). If in combination with another tense, however, 'future' is realized as 'be going to' (1985: 177). The phrase 'be going
to', as in 'I'm going to leave for ballet class now', is, therefore, a 'future' tense marker (cf. the co-presence of 'present' tense, associated with 'am'). 'Be going to' is accommodated by the following entry.

\[V.X1:Y1 \setminus \text{Fut [Perf] [Prog] [Pass]} [\text{Adv.X:Y}]\]
\[V.X1:Y1\]

(2175)

The first element, 'progressive' (Prog), consists of 'be' + 'ing' (cf.4.14 (2172)). 'Future' (Fut) is realized as 'Prog go to'. The other non-optional constituent in this entry is V.X:Y (any sub-type of verb). A verb must be present to exclude from the domain of this rule sentences with the present progressive form of 'go' such as 'I am going to ballet class now'. The optional elements listed show precisely those constituents which may intervene between the discontinuous constituents which realize 'future' or between 'future' and the main verb (Y1).

4.11 Marked word order of questions

There are five entries that realize a question element as a marked word order. The first two are related to the auxiliary 'do'. Positive questions with 'do' differ from negative questions. This is due to the involvement of 'do' in both questions and negatives. In these rules, however, we are addressing only the question element. Negatives are provided for by a different rule (cf.4.15).


(2160)

The question element (Q) on the right side of the entry
is realized as (shown on the left side of the rule) the so-called empty morpheme 'do' (Aux:do) together with a particular word order concerning tense (Tns), optional concord (C), and the noun phrase (NP). The main verb (V.Y2:X3) is included in the rule as a non-optional constituent to specify that the noun phrase (NP:X1) is functioning as the subject, not object, noun phrase. An example that illustrates the type of question provided for by this entry is: 'Did Beenie ever see you?' The rule instantiated to this example is: Tns:Past Aux:do NP:(Head: (Beenie Nmr:Sg)) Adv.Temp:ever V.T:see \ Q NP:(Head: (Beenie Nmr:Sg)) Tns:Past Adv.Temp:ever V.T:see

For the following entry, the same explanation as that for (2160) above holds, except that 'do' still appears on the right side of the rule because of the presence of the negative element 'not' which also requires 'do' (cf.4.15). (2161) is at a slightly lower level of analysis than the entry which provides for negatives. The negative (and subject verb agreement) entry is a more general rule and is therefore specified for unmarked word order (the word order specified on the right sides of the rules).


'Didn't Cecil join you' exemplifies the kind of question accommodated by (2161). The rule instantiated to this example is: Tns:Past Aux:do not NP:(Head: (Cecil Nmr:Sg)) V.T:join \ Q NP:(Head: (Cecil Nmr:Sg)) not Tns:Past
Aux:do V.T:join
The next three entries are for questions with the auxiliary 'have', the verb 'be', and the modals, respectively.

\[\text{[C:X]} \text{Tns:Y Aux:have [not]} \text{NP:X1 [Adv.Y1:X2] en } \text{V.Y2:X3} \setminus Q \text{NP:X1 [not]} \text{[C:X]} \text{Tns:Y Aux:have en } \text{[Adv.Y1:X2]} \text{V.Y2:X3}\]
\[\text{(2162)}\]

'Has Beenie been sick?' and 'Hadn't Cecil found him?' exemplify the questions provided for by this rule. The first example in rule form is: C:3-Sg Tns:Pres Aux:have NP:(Head:(Beenie Nmr:Sg)) en V.Be:be \ Q NP:(Head:(Beenie Nmr:Sg)) C:3-Sg Tns:Pres Aux:have en V.Be:be

\[\text{(2163)}\]

The part of the rule ((\text{en,ing}) V.Y2:X3, AP:Y3, NP:X4) stipulates that 'be' may be followed by either en V.Y2:X3, ing V.Y2:X3, AP:Y3, or NP:X4. These verbs and complements are present to specify that the noun phrase (NP:X1) is functioning as the subject, not complement, noun phrase.

'Is Cecil a sea serpent?' and 'Wasn't Beenie acting strangely?' are taken care of by (2163). The rule instantiated to the second example is: Tns:Past V.Be:be not NP:(Head:(Beenie Nmr:Sg)) ing V.C:act \ Q NP:(Head:(Beenie Nmr:Sg)) not Tns:Past V.Be:be ing V.C:act

\[\text{(2164)}\]

'Can Beenie come outside?', 'Couldn't Cecil swim?', and
'Should Beenie have whistled?' are examples of the type provided for by (2164). The last question represented in rule form is: Mod:should Tns:Pres NP:(Head:(Beenie Nmr:Sg)) Aux:have \ Q NP:(Head:(Beenie Nmr:Sg)) Mod:should Tns:Pres Aux:have

All of these rules are extremely sensitive. To appreciate them fully, the reader is reminded to consider them in their larger context, as part of a whole system. To see how they work together, the reader is referred to Chapter 7.

4.12 Some "hopping" (anataxis)

We want all instances of 'not' to be at the beginning of the verb phrase for the purpose of higher level entries (cf.5.6 & 5.7). 'Do', 'have', and 'be' are the elements with which 'not' may co-occur and are thus present in the following entry.

[C:X] Tns:Y (Aux:(do,have), V.Be:be) ~([not] [Adv.X1:Y1]) [ing,en] \ [not] [C:X] Tns:Y (Aux:(do,have), V.Be:be) [ing,en] [Adv.X1:Y1]

(2165)

Since 'not' sometimes intervenes between the discontinuous morphemes which are the realization of a single lexeme, this rule also ensures that the discontinuous realizations of the lexemes 'progressive', 'perfect', and 'passive' ('be' + 'ing', 'have' + 'en', and 'be' + 'en', respectively) occur next to one another (also for the purpose of higher level entries (cf.4.14)). Like 'not', adverbs may also intervene between the discontinuous constituents, so
they are included in this rule and placed at the end. (Cf. 4.18 for the rule which provides for adverbs in verb phrases.)

That 'not' and adverbs may occur in any order with respect to one another is shown by their enclosure in parentheses preceded by a tilde (cf. 2.2).

Some examples that the rule accounts for are: 'Fred has not recently eaten', 'Barney is throwing rocks', and 'Wilma does not always do the shopping'. The first example represented in rule form is: C:3-Sg Tns:Pres Aux:have not Adv.Temp:recently en \ not C:3-Sg Tns:Pres Aux:have en Adv.Temp:recently

4.13 Subject verb agreement

In present tense, verbs must agree in number with third person singular subjects.

\[
\text{NP:(..Head:(X Nmr:Sg)\ldots) [not] [Adv.Y:X1] C:3-Sg Tns:Pres \ NP:(..Head:(X1 Nmr:Sg)\ldots) [not] Tns:Pres [Adv.Y:X1] IF X NOT = (1, 2) } \tag{2166}
\]

Agreement is specified by the presence of singular number (Nmr:Sg) on the head of the noun phrase (NP) and the concord (C) element third person singular (3-Sg). There are only two exceptions to this general rule, which are provided for by the IF statement at the end of the entry. The head of the noun phrase (X) may not be 'I' or 'you'. These are specified as '1' and '2', respectively (cf. 3.3).

'Fred works in a quarry' and 'He operates a dragline' are examples which reflect the subject verb agreement provided for by this rule.
Third person singular subjects also agree with the verb 'be' in past tense.

\[ \text{NP:} (\ldots \text{Head:} (X \text{ Nmr:Sg}) \ldots) [\text{not}] \text{C:} 3\text{-Sg} \text{ Tns:Past V.Be:be} \]
\[ \quad \text{IF } X \text{ NOT } = (1, 2) \]  
\[ (2169) \]

As in (2166) above, the elements '1' and '2' are excluded from the rule by the IF statement. Any other singular noun or pronoun (X Nmr:Sg) and the concord element 3-Sg, in conjunction, specify agreement.

Examples taken care of by this rule are: 'Was the man a carpenter' and 'The man was a lumberjack'. The rule instantiated to the latter example is:

\[ \text{NP:} (\text{the Head:} (\text{man Nmr:Sg})) \text{ C:} 3\text{-Sg} \text{ Tns:Past V.Be:be} \quad \text{NP:} (\text{the Head:} (\text{man Nmr:Sg})) \text{ Tns:Past V.Be:be} \]

The two subsequent entries concern agreement of first person singular subjects with the verb 'be'.

\[ \text{Q NP:} (\text{Head:} (1 \text{ Nmr:Sg}) \ldots) \text{ not } [\text{C:} 1\text{-Sg}] \text{ Tns:Pres V.Be:be} \]
\[ \quad \text{Q NP:} (\text{Head:} (1 \text{ Nmr:Sg}) \ldots) \text{ not } \text{Tns:Pres V.Be:be} \]  
\[ (2167) \]

If the three elements 'Q', 'not', and 'Pres' are all present on the content side of the rule, then there are alternate realizations, one with the concord element 1-Sg, one without it. These options enable such phrases as 'Aren't I ...' and 'Am I not ...'. The former example illustrates the alternant without the concord element (cf. 'are' (1405) and 3.8), the latter with the concord element (cf. 'am' (1404) and 3.8).

The following entry is the default rule for first person singular subject verb agreement.
In contrast to (2167), the tense in (2168) may not only be 'present', as in 'I am in the choir', but also 'past', 'I was in the choir', since tense is represented by the variable X. 'Not' is an optional element: contrast 'I am not in the choir' with the examples above.

4.14 Aspect and Voice

Two kinds of aspect, progressive and perfect, and passive voice are realized in English as discontinuous morphemes.

The lexeme 'progressive' is realized as the verb 'be' and the present participle element 'ing', as in 'She is singing loudly' (cf.3.8 (1403,1986)).

V.Be:be ing \ Prog (2172)

The 'perfect' lexeme is realized as the auxiliary 'have' and the past participle element 'en', as in 'She has eaten all the cookies' (cf.3.8 (1416,1462,1984,1985)).

Aux:have en \ Perf (2171)

The 'passive' lexeme is realized as the verb 'be' together with 'en', as in 'She was eaten by tigers' (cf.3.8 (1406,1462,1984,1985)).

V.Be:be en \ Pass (2173)

4.15 Negative

'Not' may occur only if accompanied by a modal, the auxiliary 'have' or 'do', or the verb 'be'. If no such
element is present, expressed by the zero (Ø) option at the end of the content side of the rule, then part of the realization of 'Neg' is the auxiliary 'do'.

\[ NP:X (\text{Mod}:Y \text{not Tns}:X_1, \text{not Tns}:X_1 (V.\text{Be}:be, \text{Aux}:\text{(have,do)})) \setminus NP:X (\text{Neg Mod}:Y Tns:X_1, \text{Neg Tns}:X_1 (V.\text{Be}:be, \text{Aux}:\text{(have,Ø)})) \] (2170)

The noun phrase (NP) is needed in this entry to specify that this rule provides for unmarked word order (cf.4.11 (2164)). 'We have not gone on stage yet', 'Your jump is not high enough', 'They do not ever put on enough lipstick', and 'They cannot come to the performance' are covered by this rule. The third example, represented in rule form, has no auxiliary or 'be' on the content side. 'Neg' is, therefore, realized as 'not' and 'Aux:do'.

\[ NP:(\text{Head}: (3 \text{ Nmr}:P_1)) \text{not Tns}:\text{Pres} \text{Aux}:\text{do} \setminus NP:(\text{Head}: (3 \text{ Nmr}:P_1)) \text{Neg Tns}:\text{Pres} \]

4.16 Verbals

Verbals are parts of verb phrases and may consist of one of the following: an intransitive verb, a transitive verb and object noun phrase, a passive element together with a transitive verb, or a copula and complement.

The first type, with intransitive verbs, is accommodated by this entry.

\[ V.I:X \setminus Vbl:(V.I:X) \] (2200)

(2200) allows for the intransitive verbs in such statements as 'Fred slept' or 'Wilma sat by the window' (cf.3.8 (1557,1555)).

Transitive verbs and their objects are covered in the
next entry.

\[ V.T:X \ NP: (. . Head: (Y Nmr:X1) . . ) (\emptyset, m \ IF \ Y = (1, 2, 3, 3-Y1, who)) \ \ Vbl: (V.T:X \ NP: (. . Head: (Y Nmr:X1) . . )) \]

'(Wilma made rock candy' is a sentence whose verbal is accommodated by this rule. The noun phrase (in this case 'rock candy'), following a transitive verb, is functioning as the object of the verb. Noun phrases with nouns as head are unmarked for objective case, hence the '∅'. If a pronoun is the head of the noun phrase, as, for instance, in 'Wilma made them', then following a transitive verb, it must show objective case marking, hence the 'm' (cf.3.3 (1113)).

These two examples represented in rule form are,
respectively:

\[ V.T: make \ NP: (Modif: rock \ Head: (candy Nmr:Sg)) \ \ Vbl: (V.T: make \ NP: (Modif: rock \ Head: (candy Nmr:Sg))) \]

\[ V.T: make \ NP: (Head: (3 Nmr:P1)) \ m \ \ Vbl: (V.T: make \ NP: (Head: (3 Nmr:P1))) \]

The passive element (cf.4.14) and the accompanying transitive verb (cf.3.8) are provided for in the following rule.

\[ \text{Pass} [\text{Adv.X:Y}] \ V.T: X1 \ \ [\text{Adv.X:Y}] \ Vbl: (\text{Pass} \ V.T: X1) \ \ \ \text{IF} \ X \ \text{NOT} = \ \text{Loc} \]

(2202)

The adverb that occurs in this verbal contraction is only minimally specified. Locative adverbs such as 'here' and 'there' (cf.3.6 (1800-1803)) do not usually precede a verb, for example, *'The woman was there seen'.

'The candy was eaten by Fred' illustrates this kind of verbal. The rule instantiated to this example is:
Pass V.T:eat \ Vbl:(Pass V.T:eat)

Certain types of verbs may take adjectives as complements. 'Be' and the simple copulas (V.C) (cf.3.8 (1440-1449)) are such verb types.


'Fred is fat' ('be') and 'Wilma looks pretty' (copula) are examples of this type verbal.

The verb 'be', the complex copulas (V.CCx) (cf.3.8, 4.20 (2001)), and the (simple) copula 'become' can take a noun phrase as complement (cf.3.8).


Some example sentences provided for by (2204) are:
Fred became a father', 'Wilma seems like a good mother', and 'Fred is a quarryman'.

Predicate nominatives, as these types of complements are called, must agree in number with their antecedents.

The following rule provides for that agreement.

NP:((..Head:(X Nmr:Y)...) .. Vbl:((..(V.Be:be, V.C:become, V.CCx:X1) NP:((..Head:((Y1 Nmr:Y, (who,which,what) Nmr:Sg))) \ NP:((..Head:((X Nmr:Y)...) .. Vbl:((..(V.Be:be, V.C:become, V.CCx:X1) NP:((..Head:((Y1,who,which, what) Nmr:Y))) (2502)

The instances of the variable Y in this rule signify the number agreement between the predicate nominative and its antecedent. This check prohibits such combinations as *
'They are a doctor', and allows 'They are doctors', as in:
NP:(Head:3 Nmr:Pl)) Tns:Pres Vbl:(V.Be:be NP:(Head:(doctor
Nmr:Pl))) \ NP:(Head:3 Nmr:Pl)) Tns:Pres Vbl:(V.Be:be
NP:(Head:(doctor Nmr:Pl)))

'Who', 'which', and 'what' are singular (cf.4.7 (2155))
and may not agree with the antecedent. They are accommo-
dated by the rule nonetheless and interpreted as having the
plurality of their antecedent.

4.17 Participial phrases

The function of the participial phrase as a noun
attribute is discussed above (cf.4.8). Its internal
structure is described by the following entry.

ing [Perf] Vbl:X \ PartP:(ing [Perf] Vbl:X) \hspace{1cm} (2250)

A participial phrase consists of the present participle
element 'ing' followed by an optional 'perfect' and a
verbal. 'Dancing' (ing Vbl:(V.I:dance)), 'having eaten the
pie' (ing Perf Vbl:(V.T:eat NP:(the Head:(pie Nmr:Sg)))),
'having been reduced' (ing Perf Vbl:(Pass V.T:reduce)),
'being sick' (ing Vbl:(V.Be:be AP:(Adj.Subj:sick))), and
'looking like the devil' (ing Vbl:(V.CCx:look-like NP:(the
Head:(devil Nmr:Sg)))) are examples of participial phrases
provided for by this statement.

4.18 Verb phrases

The term "verb phrase" is used above in a general
sense; here, it is defined by a specific construction.

[[[Perf] [Fut [Perf]]] [Prog] [Adv.X:Y] Vbl:X1 \ 
VP:([[[Perf] [Fut [Perf]]] [Prog] [Adv.X:Y] Vbl:X1)
IF X NOT = Loc \hspace{1cm} (2300)

The embedded square brackets show the allowable combi-
nations of 'perfect' (Perf), 'future' (Fut), and 'progressive' (Prog). Notably, for the second instance of 'perfect' to occur, 'future' must also be present (cf. 2.2). The analysis of these elements is based on that given by Halliday (1985: 180-1) (cf. 5.3).

The optional aspectual elements, 'progressive' and 'perfect', and one of the tense elements 'future' (from 'be going to' (cf. 4.10 (2175)), together with an optional pre-verbal adverb and a verbal occur as a verb phrase. 'Has been promoted' (Perf + Vbl), 'is looking good' (Prog + Vbl), and an example of the longest possible verb phrase 'has been going to have been being completely held back' (Perf + Fut + Perf + Prog + Adv + Vbl) are all provided for by (2300). As noted above (cf. 4.16), pre-verbal adverbs are not usually locative, expressed by their exclusion in the IF statement.

4.19 **Predicates**

Predicates include the verb phrase (VP) (cf. 4.18), negative (Neg), tense (Tns), and modal (Mod) elements, and all adverbs (Adv) and prepositional phrases (PP) which follow the verb phrase.

\[
\begin{align*}
[Neg] & \ [Mod:X] \ Tns:Y \ VP:X1 \ ~ ([Adv.Y1:X2] \ [PP:Y2])^* \\
& \ Pred: ([Neg] \ [Mod:X] \ Tns:Y \ VP:X1 \ ~ ([Adv.Y1:X2] \ [PP:Y2])^*)
\end{align*}
\]

\(2400\)

There may be present both adverbs and prepositional phrases, and they may be in any order with respect to one another (cf. 2.2); for example, 'The car hit me in the rear end there by the license plate'.
The prepositional phrase at the end of the predicate functions adverbially. It may be locative such as 'on the table', temporal like 'on Christmas', manner as in 'in a similar way', or probability such as 'for sure'. The prepositional phrase can also be the by-phrase of a passive, for example, 'by the President' in 'The order was issued by the President'. The types of prepositional phrases are resolved in the Semicon where lexemic and sememic information interface.

4.20 Other phrase structure entries

In this section, entries which have a variable that has a limited number of instantiations are discussed. Two notable examples are the hour noun (N.Hour) 'X-o'clock', in which X can be only 'one' through 'twelve', and complex copulas (V.C.Cx) 'X-like', in which X may apply to only about fifteen verbs in the language.

\[
\begin{align*}
\text{Num.Card:X} & \quad \text{Nmr:Y} & \quad \text{Adv.Temp.Hr:o'clock} & \quad \backslash \\
\text{N.Hour:X-o'clock} & \quad \text{IF X} = \text{one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve} & \quad (2033) \\
\text{V.C:X} & \quad \text{Prep:like} & \quad \backslash \quad \text{V.C.Cx:X-like} & \quad (2001)
\end{align*}
\]

The latter entry provides for any of the (simple) copulas (V.C) (cf. 3.8 (1440-1449)) followed by the preposition 'like'. Together, they are the realization of a complex copula (V.C.Cx). For example, the (simple) copula 'look', as in 'He looks sick', when followed by 'like' is the complex copula 'look-like', as in 'He looks like a nut'. Co-occurrence restrictions in phrases (cf. 4.16) give rise to the different tactic classification of these complex forms.
Some examples accommodated by (2033) are: 'Ten o'clock is a good time for me' and 'Be home by twelve o'clock'.

Some examples provided for by (2001) are: 'We became like brothers' and 'It sounds like a good idea'.
NOTES

Chapter 4

1 An alternative approach in the writing of entries calls for every tactic label to reflect all of the information contained within the phrase dominated by the label, if that information is used in restricting the occurrence of the phrase. For example, the adjective phrase label (AP) in the adjective phrase rule, would have (minimally), in this alternative approach, a suffix showing the sub-type of adjective which is its head element (cf. AP.Y1, below). The rule would, therefore, look like this:


These two alternatives have virtually the same information. With the simple matching strategy envisioned for the processing of the linguistic information housed in these entries (cf.2.4), either type of rule works equally well.

2 Probably locatives, temporals, probability, and frequency adverbs will be found to bind with the verb, while manner and degree adverbs will bind with the adjective.

3 Actually, we don't want this to mark tense for ",any" verb, only the finite verb. The following amendment to the rule constrains it somewhat.

Tns:(X,Ø) [Y] (Aux,V.X1):Y1 \ Tns:(X,Pres) [Y]
(Aux,V.X1):Y1 IF Y NOT = (en,ing)

In an example such as 'I have seen the cat', 'have seen' is analyzed as 'Aux:have en V.T:see'. The amended rule assigns Tns:Pres to Aux:have but not to V.T:see since the latter is preceded by 'en' (Y in the rule).

Unfortunately, this amendment isn't enough because (the second instance of) 'be' in 'I am going to be a dancer' (..V.I:go Prep:to V.Be:be) has no tense (Tns:Ø) and no 'en' or 'ing' preceding it, yet it should not have a tense element assigned to it. Until this rule can be constrained to apply only to finite forms of the verb, we hedge and make a general rule for the decoding process which uses (2110) that says this rule can apply only once per clause.

4 The scope of "verbal", "verb phrase", and "predicate", as used in this thesis, is based on class notes from Lamb, Rice University, 1986.

5 Actually, this rule prohibits these undesirable combina-
tions in the encoding process. In decoding, however, this rule simply would not match and therefore not apply to a string represented by 'They are a doctor', but that string would be decoded nonetheless. This is desirable since we want to understand things that we do not necessarily want to produce. (Cf. colloquial examples such as 'They are a cute couple'.)
CHAPTER 5
THE SEMICON

5.0 Entries of the Semicon

The Semicon has two main kinds of information in it, semantic relations and sememes. The latter are presented first in the discussion (cf. 5.1) since the properties of their intensional definition figure prominently in restricting the co-occurrence of elements in propositions. Semantic relations, the interface between the lexemic and sememic levels, between phrase structure and clause structure, is then presented in 5.2 - 5.7.

5.1 Sememes

In 1.2, the intensional definition of sememes is discussed. The relations of the sememe to its intension are predicative, propositional, and pragmatic. Examples of these relations are 'a cat is an animal' (predicative), 'a cat eats meat' (propositional), and 'don't say the word cat around an ailurophobe' (pragmatic). The properties which constitute the intension of each sememe can be specified in the semiotic format in the manner shown by the following entries.

\[ \text{man} \ \backslash \ \text{human} + \text{male} + \text{adult} + \text{etc.} \quad (3500) \]

\[ \text{Houston} \ \backslash \ \text{place} + \text{city} + \text{in Texas} + \text{etc.} \quad (3501) \]

On the content side of the rule, the properties are joined by a plus sign (+) (cf. 2.2). Wherever all of the
properties which conjoin to yield a sememe are not specifically stated, 'etc.' is used to indicate their presence.

The predicative relation, a hyponymic relationship where the hyponym is a "kind of" the hypernym, allows each sememe to fit into a taxonomy. 'Thing' is at the root of the taxonomy for objects (as opposed to actions). The following entry states that 'a concrete object is a thing'. (Of course, this 'thing' is differentiated from the morpheme/lexeme/sememe 'thing' which appears on the LEFT side of an entry, cf. thing \ thing.)

concrete-object \ thing

Although this statement captures the relationship of 'concrete-object' to 'thing', further information about what concrete means must be provided in order to differentiate 'concrete-object' from 'abstract-object' which is also a 'thing'. Rule (3600), shown in part above, has another property, 'tangible'.

concrete-object \ thing + (potentially) tangible

The hedge 'potentially' is inserted here to allow for such things as unicorns and quarks.

'Concrete-object' is a concept that has two sub-cATEGORIES, 'animate' and 'inanimate':

animate \ concrete-object + living

inanimate \ concrete-object

'Animate', in turn, is a hypernym of 'animal' and 'plant':
animal \ animate + etc. \ (3605)
plant \ animate' + etc. \ (3606)

'Human' is subsumed under the more general class 'animal':

human \ animal + etc. \ (3607)

In entries of these types, the essential information that is needed to conclude that 'man' is a 'thing' is present. Part of the intension of 'man', (i.e., one of its sememic properties) is 'human', which is a kind of 'animal', which is a kind of 'animate' object, which is a kind of 'concrete-object', which is a kind of 'thing'.

By using the hierarchical structure established in these entries, that which is referred to in the negative (minus X) can also be discriminated. For example, if an entry states that an element has the sememic property 'minus animate' (specified in an entry as $\neg$-animate$)$, then that element does not have a link in the system to the concept 'animate'. In hierarchical terms, it is not dominated by the node labelled 'animate'.

There are many types of oppositeness in language: complementarity in which there are only two possibilities, such as 'male:female' or 'dead:alive'; antonymy in which the opposition is between two points on a graded scale, such as 'high:low' or 'short:tall'; and converseness, such as 'buy:sell' or 'do:undo', etc.

The only type of oppositeness that is addressed in the present grammar is antonymy (cf.5.2). Antonymy is being used here in the restricted sense in which Cruse (1986,
after Lyons 1968) uses it. It is not '... a cover term for all types of lexical opposite. Antonymy is exemplified by such pairs as long:short, fast:slow, easy:difficult, good:bad, hot:cold.' Most antonyms are adjectives whose 'members ... denote degrees of some variable property ...' (Cruse 1986:204).

The representation of antonymy could be handled in any of several ways. Cruse states that '...opposites typically differ along only one dimension of meaning: in respect to all other features they are identical...' (1986:197). If the intensional definition of each sememe were fully specified in the grammar, then antonyms could be distinguished by the overlap of all but one feature in their meanings. The meaning of each sememe has not been fully specified, however, so another method is used to stipulate antonymy. The antonym is simply indicated along with the meaning of the concept, in much the same way that dictionaries do it, for example:

\[
\text{old} \ \ \text{age + etc.} \ \ \text{ant:} \ (\text{new,young}) \ \ \ (3650)
\]

The antonym (ant) of 'old' is either 'new' or 'young'.

Event sememes are defined in the conceptual network in much the same way as 'things'; that is, they may be viewed as sub-classes of more general categories such as 'do' or 'be', distinguished, of course, by their connections to other sememes. A taxonomy of events is thus established. (See Ikegami 1969 for a discussion of semological structure of events.)
It is not our intention to present an exhaustive
treatment of either the taxonomic or propositional structure
of event sememes, but rather to offer the reader an idea of
what such an entry looks like. The following examples
express in dictionary notation a portion of a relational
network diagram for 'walk' from Ikegami (1969:108).

\begin{verbatim}
walk \ go + (step + at the same time as) \ (3003)
\end{verbatim}

According to Lockwood (1972:142), 'event sememes are
divided into classes according to the number and type of
participants which may accompany them. Also, there are
various circumstantial attributes ... such as time, loca-
tion, and manner.' These participants and circumstantial
presupposed by an event are the requirements sometimes
called event valency. The valency of an event is part of
its intension.

The required participants and/or circumstantialssuch
as time, location, and manner, are indicated by symbols
enclosed in angle brackets in the entry for the sememe.
'Eat', 'put', and 'walk', for example, have these require-
ments in addition to other properties which are expressed in
the following simply as 'etc'.

\begin{verbatim}
eat \ etc. (<animate> <concrete-object>) \ (3001)
put \ etc. (<animal> <thing> <locative>) \ (3002)
walk \ etc. <animal> \ (3003)
\end{verbatim}

'Eat' requires two participants, the eater and the
eatee. The eater must be animate and the eatee a concrete-
object. 'Put' requires three participants: someone puts
something somewhere. 'Walk' requires a single participant, as in 'The baby is now walking'.

The sememic properties of prepositions given below are based on a componential analysis of prepositions by Bennett (1975:92-93 and 1970:265).

<table>
<thead>
<tr>
<th>Preposition</th>
<th>Sememe</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>at \ locative</td>
<td>(3750)</td>
<td>(3750)</td>
</tr>
<tr>
<td>on \ locative + surface</td>
<td>(3751)</td>
<td>(3751)</td>
</tr>
<tr>
<td>in \ locative + interior</td>
<td>(3752)</td>
<td>(3752)</td>
</tr>
<tr>
<td>by \ end point in time</td>
<td>(3755)</td>
<td>(3755)</td>
</tr>
<tr>
<td>by \ locative + proximity</td>
<td>(3753)</td>
<td>(3753)</td>
</tr>
<tr>
<td>to \ goal</td>
<td>(3754)</td>
<td>(3754)</td>
</tr>
</tbody>
</table>

Preposition sememes require certain kinds of participants just as event sememes do. Unlike event sememes which presuppose (perhaps) several participants, however, prepositions presuppose only one (their so-called object). The preposition and its participant together function as a locative, temporal, or manner circumstantial, which, in turn, is presupposed by event sememes (cf.(3002) above). For example, 'at' requires a participant that is either a 'temporal-point', such as 'X-o'clock' (cf.3503) or a 'place', such as 'the beach'. The whole phrase 'at the beach' has a functional meaning of 'locative' and, as such, is required by certain events such as 'put' or 'stay'.

Below are the same set of preposition sememes given above but with their required participant types enumerated in angle brackets.

<table>
<thead>
<tr>
<th>Preposition</th>
<th>Sememe</th>
</tr>
</thead>
<tbody>
<tr>
<td>at \ locative &amp; (temporal-point, place)</td>
<td>(3750)</td>
</tr>
<tr>
<td>on \ locative + surface &amp; (concrete-object, day)</td>
<td>(3751)</td>
</tr>
</tbody>
</table>
in \ locative + interior <(concrete-object, abstract-object IF NOT (day Sg) AND temporal-point)>

(3752)

by \ locative + proximity <concrete-object>

(3753)

by \ end point in time <temporal>

(3755)

to \ goal <(concrete-object,temporal)>

(3754)

'Temporal' refers to three sub-categories: 'temporal-point', 'day', and 'time-period' (cf.3780-3782). 'Temporal-points' (cf.3780) are points in time, for instance 'X-o'clock' (cf.3503). 'Beenie got up at eight o'clock' illustrates this type of participant. 'Day' (cf.3782) characterizes the days of the week like 'Monday' (cf.3502) and holiday names like 'Christmas' (cf.3551) and 'Easter'. 'Cecil came on Monday' exemplifies this participant type and it co-occurrence with 'on'. 'Time-period' (cf.3781) includes 'year', 'month', 'week', names of months such as 'February' (cf.3552), names of seasons like 'summer', names of years like '1988', and general times of day such as 'evening', 'morning', and 'hour'. 'By February, Cecil was back' and 'Beenie slept from winter to summer' illustrate this participant type.

Some of the other sememes and their intensional definitions represented in the Semicon are:

now \ temporal + etc. (3558)
here \ locative + etc. (3555)
black \ quality + color + etc. (3651)
the \ definite (3675)
that \ definite + distal (3676)
this \ definite + proximal (3677)
a \ indefinite (3680)
(P1, plural) \ more than one (3700)
(Sg, singular) \ one (3701)
who \ human + etc. (3685)
which \ -human + etc. \ (3686)
that \ thing + etc. \ (3687)

The reader will notice that 'that' is listed in the Semon twice (cf.3676,3687). They are placed in separate statements because the element on the expression side of the entry represents a sememe, a concept, and these two concepts do not overlap at all. In the Morphicon, the graphemic string for 'that' is listed only once (cf.1153), but the different morphemes are kept separate by different tactic codes (cf.3.3).

5.2 \ Lexotactic constraints and sememic properties of participants

This section deals with some of the more salient co-occurrence restriction problems. The treatment of these issues is not exhaustive, but what is shown may serve as a model for the handling of other, similar problems.

Agreement between the relative pronoun 'who' or 'which' and its antecedent depends on the sememic property 'human'. For example, in 'the man who ate the cake', 'man' has the sememic property 'human' (cf.5.1 (3500)), and so does 'who' (cf.5.1 (3685)). They therefore agree. In 'the story which I told', 'story' has the sememic property 'abstract-object' (cf.3504), which is both on a different branch and higher in the taxonomy of 'things' than 'human', and therefore not 'human' (cf.5.1). 'Which' has the property '-human' (minus human) (cf.5.1 (3686)), so 'story' and 'which' agree. The relative pronoun 'that' does not discriminate humanness (cf.5.1 (3687)) and so can be used to refer to either humans
or nonhumans, for example, 'the man that ate the cake' and 'the story that I told'.

The following rule states that the pronoun lexeme in the relative clause construction (cf. X1 on the left side of the rule) is properly specified if the sememe which is its realizate (cf. X1 on the right side of the rule) has the same sememic property (Y1) as its antecedent (X).

NP: (. . Head: (X Nmr: Y)) [m] Relv . . NP: (Head: (X1 Nmr: Y))
\ NP: (. . Head: (X $Y1$ Nmr: Y)) [m] Relv . . NP:
  (Head: (X1 $Y1$ Nmr: Y)) (3815)

The sememic property is specified between two dollar signs to the right of the sememe whose property it is (cf. 2.2). Agreement between the antecedent and the relative pronoun is indicated by the same variable in each of these positions.

'm' is present in the rule to provide for a pronoun antecedent which is functioning as an object (cf. 4.9, 4.16), as with 'him' in 'I gave it to him, who gave it to her'.

The first example above, 'the man who ate the cake', represented in rule form, is:

NP: (the Head: (man Nmr: Sg)) Relv NP: (Head: (who Nmr: Sg)) \ NP: (the Head: (man $human$ Nmr: Sg)) Relv NP: (Head: (who $human$ Nmr: Sg))

The instantiation of the variable for the sememic property, indicated as Y1 in the rule, comes from the entries for the sememes whose property it is, in this case the entries for 'man' and 'who' (cf. (3500, 3685)).

The following entry stipulates that noun phrases (NP)
are the lower level manifestations of participants (P). The label noun attribute (Nattr) is interpreted at the semantic level as a participant attribute (Pattr).

\[
\text{NP:} (.\text{Head:} (X \text{ Nmr:} Y) [\text{Nattr.} X1: Y1]^*) [X2] \ \\
\text{P:} (.\text{Head:} (X Y) [\text{Pattr.} X1: Y1]^*) [X2] \text{ IF } X2 \text{ NOT} \ = \text{ Relv} \quad (3925)
\]

Noun phrases followed by 'Relv' are excluded from the domain of this rule by the IF statement. The element 'Relv' is present at a level of analysis lower than the one indicated in this entry. When elements in a relative clause are being processed (cf.7.2), the element 'Relv' is present until the clause is constructed. Then, the tactic label RelvCl is assigned (cf.5.7 (3991)), and the whole clause is rank-shifted (cf.4.8) to the phrase (lexemic) level where it is interpreted as a noun attribute (cf.4.8 (2600)). The IF statement prevents the application of (3925) during processing until the relative clause has been analyzed as part of the noun phrase.

The following rule treats present participial phrases as reduced (rank-shifted) clauses (after Quirk and Greenbaum 1975:376 (c), and Winograd 1983:240-1,476).

\[
\text{P:} (.\text{Head:} (X Y) .\text{Pattr.} \text{PtP:} (\text{ing [Perf]} \text{ Vbl:} X1)) \ \\
\text{P:} (.\text{Head:} (X Y) .\text{Pattr.} \text{PtP:} (\text{P:} (\text{Head:} (X Y)) \ \\
\text{Pred:} (\text{Time:} (\text{Pres [Past]} \text{ Vbl:} X1))) \quad (3940)
\]

On the left side of the entry is a participant (P) whose attribute is a participial phrase (Pattr.PtP). On the right side of the entry is a participant whose attribute is a clause. (A clause consists of a participant (P) functioning as subject, followed by a predicate (Pred) (cf.
5.6).)

On the right side of the rule, the (subject) participant in the clause (cf. P: (Head: (X Y)) at end of second line of rule) is the same as the participant that the phrase cum clause is modifying (cf. P: (Head: (X Y)) at beginning of second line of rule). For example, in 'the woman wearing the dress', since 'the woman' is the one doing the 'wearing', she functions as the subject of the clause, as the following abbreviated representation of the instantiated rule shows: P: (the woman Pattr.PtP: (wearing the dress)) \ P: (the woman Pattr.PtP: (P: (the woman) Pred: (is wearing the dress))) The internal integrity of the participant at the lower level is present at the higher level since the head of the participant is stated in two places.

The non-optional parts of a predicate, specified on the right side of the entry, are time (Time) and verbal (Vbl) (cf.4.19, 5.6). The verbal is present at both levels in (394G). 'ing' is the realization of 'present' (Pres) time, and 'perfect' (Perf) is the realization of past time in combination with another time (in this case, present time) (cf.5.3 (3800)).

The following rule specifies that adjectives in the same phrase must be sememically compatible. In 4.4, it is mentioned that the grammar at the lexemic level allows antonyms to co-occur. This entry restricts their co-occurrence.

Two adjectives of the same sub-type (indicated by Adj.X) may not occur in the same adjective phrase (AP,Modif) if the antonym (ant) of the first adjective is the second adjective (indicated by the IF statement). For example, in '?the hot, cold water', 'hot' and 'cold' are both subjective (Subj) adjectives (Adj), and they are antonyms. The attempt during processing to instantiate (3820) to 'hot, cold' is prevented by virtue of the following rule in the Semicon: hot \ etc. ant:cold, which causes the IF statement of (3820) to look like: IF cold NOT = cold. Such an incongruity would trigger, during processing, a response by the system equivalent to the human response "huh?"

The antonym is shown between dollar signs since it is in the dictionary as part of the intensional information (cf.5.1). It is in square brackets, indicating optionality, because not all adjectives have lexical opposities (cf. 'red', 'wooden').

The following statements provide for the reconciliation of tactic labels and sememic properties within predicate adjectives (AP,PrAdj) and modifiers (Modif,PModif).

(AP,Modif): (...)Adj.Subj:X) \n(PrAdj,PModif): (.X $subjective$)

(AP,Modif): (...)Adj.Obj:X) \n(PrAdj,PModif): (.X $size,shape$)

(AP,Modif): (...)Adj.Age:X) \n(PrAdj,PModif): (.X $age$)
(AP,Modif):(Adj.Col:X) \n   (PrAdj,PModif):(..X $color$)  (3828)

(AP,Modif):(..Adj.Clas:X) \n   (PrAdj,PModif):(..X $classifier$)  (3829)

The sememic properties 'age', 'color', 'size' and
'shape', shown in the above entries between dollar signs,
are present in the entries of each adjective sememe of that
type, for example:

    old \ age  (3650)
    black \ color (3651)

The sememic properties 'subjective' and 'classifier' (cf.
(3825,3829)) are not present in any entries for adjective
sememes, however. Rather, there are sememic properties of
adjectives classified at the lower of the two levels as
Adj.Subj and Adj.Clas, which, in turn, have the sememic
property 'subjective' or 'classifier'. For example, part of
the meaning of an adjective such as 'sad' (cf.3652) is that
it is an 'emotion'. 'Emotions', in turn, have a property
'subjective' (cf.3609). Other 'subjectives' are 'humor',
'opinion', and 'sensory perception'. By following an
indirect path through a taxonomic hierarchy such as the one
discussed in 5.1, certain adjective sememes may thus be
interpreted as having the property 'subjective'. For
example, X, in (3825), instantiated to 'sad', as in 'The sad
boy cried', is:

Modif:(Adj.Subj:sad) \ PModif:(sad $subjective$)

X can be instantiated to 'sad' due to the presence of these
entries:
sad \ emotion + etc. (3652)
emotion \ subjective + etc. (3609)

X, in (3827), instantiated to 'old', as in 'The car is old', is:
AP:(Adj.Subj:old) \ PrAdj:(old $age$)

X can be instantiated to 'old' due to the presence of entry (3650) shown above.

The lexotactic labels AP and Modif and the semotactic labels PrAdj and PModif are present to constrain the domain of the rule. In encoding, for example, only elements which are functioning as predicate adjectives (PrAdj) or participant modifiers (PModif) should be analyzed as "Adj.X", not just any sememe with the specified property. For example, 'emotion' has the property 'subjective' (cf.3609,above), but Adj.Subj is an inappropriate label for 'emotion'. 'Emotion' should be realized as a noun, not an adjective.

The following entry takes care of the reconciliation of tactic labels for and sememic properties of all the other elements within adjective phrases.

\[(PrAdj,PModif):(([X:Y,Intns:X1] Y1)*) \ \ (PrAdj,PModif):([Y,X1 $intensity$] Y1)*)\]
\[IF \ X \ NOT = \ Adv.X2 \]

(3830)

Y1 is the adjective sememe (cf. the right side of entries (3825-3829) above).

X1 is an intensifier such as 'very' (cf.3.6). The reason for reconciling the tactic label intensifier (Intns) and the sememic property 'intensity' in this particular rule is two-fold. First, X1, on the right side of the rule, must have a tactic label such as PrAdj or PModif to distinguish
it from other elements with different tactic labels but with the same sememic property. (Cf. a 'thing' such as 'loudness' has the property 'intensity', but intensifier is not an appropriate lextotactic label for 'loudness' which should be realized as a noun.) Second, the reconciliation is done in the adjective phrase entry since that is the only construction in which intensifiers occur in the present grammar.

Y in (3830) is a locative, temporal, or other circumstantial expression, whose tactic label and sememic property reconciliation is provided for by rules given in 5.4. The IF statement is present to specify the level of analysis of the circumstantials in the rule, preventing the application of this rule before those in 5.4 during processing.

Pronouns have certain characteristics of the nouns which they represent. In the following entry, a pronoun has the same plurality, represented by X on both sides of the entry, as its referent.

\[
\text{Head:}((1,2) \ X) \ \ \text{Proform:}((\text{Head:}(\text{Speaker},\text{Hearer}) \ X))
\]

(3810)

'1' and '2' have been defined elsewhere as first person (cf.3.3 (1100,1101,1107,1108)) and second person (cf.3.3 (1102)), respectively. 'Speaker' and 'Hearer' are specific concepts denoting the interlocutors. 'Proform' is a semotactic code which indicates that the 'thing' following it, in this case, Speaker or Hearer, is to be referred to in its pronominal form.
The rule instantiated to 'I' is:  \[\text{Head:}(1 \text{ Sg}) \ \backslash \ \text{Proform:}(\text{Head:}(\text{Speaker Sg}))\]. 'I', or '1' as it is symbolized in the system, is the pronominal form (proform) of a single 'Speaker'.

Number agreement is necessary between a pronoun and its referent even if the pronoun refers to something other than the 'Speaker' or 'Hearer'.

\[
\text{Head:}(3 \text{ Pl}) \ \backslash \ \text{Proform:}(\text{Head:}(X \text{ Pl})) \ \text{IF} \ X \ \text{NOT} = (\text{Speaker, Hearer}) \quad (3812)
\]

For example, to refer to 'cats' in its pronominal form (i.e., Proform:((Head:(cat Pl)))(3 Pl) (i.e., 'they' (cf. 1106) or 'them' (cf.1113)) is the representation to use.

The IF statement must be present to specify that the 'Speakers' or 'Hearers', which could be the instantiation of the variable X, may not be realized as (3 Pl). (Cf. in encoding, we want the representation Proform:((Head:(Speaker Pl)) to use rule (3810), not rule (3812).)

If the 'thing' being referred to is singular and not the 'Speaker' or 'Hearer', then agreement of gender as well as number is required.

\[
\text{Head:}(3-X \text{ Sg}) \ \backslash \ \text{Proform:}(\text{Head:}(Y \text{ Sg})) \ \text{IF} \ Y \ \text{NOT} = (\text{Speaker,Hearer}) \quad (3811)
\]

Gender is indicated in the rule by the variable X. The pronominal form of 'man', 'he' or 'him' (cf.1103,1110), and the concept 'man' (cf.3500) have, for example, the same sememic property 'male', shown in rule form as:

\[
\text{Head:}(3-\text{male Sg}) \ \backslash \ \text{Proform:}(\text{Head:}(\text{man Sg}))
\]
5.3 **Aspect, tense, and time**

We follow Halliday's analysis of progressive and perfect aspect as secondary tenses (1985:177, 180-1). The primary tense, represented in the present grammar as tense (Tns), indicates the time of the action '...relative to the speech event'. The secondary tenses indicate the time of the action '...relative to the time selected in the previous tense.' (Halliday 1985:177) 'Perfect' (Perf) is a secondary 'past' tense, and 'progressive' (Prog) is a secondary 'present' (Pres) tense. The primary and secondary tenses are represented at the sememic level as time (Time).

\[
\text{Tns: } X \ VP: ([\text{[Perf]} \ [Fut \ [Perf]]] \ [Prog] ..) \ \backslash \ \\
\text{Time: } (X \ [[Past] \ [Fut \ [Past]]] \ [Pres]) ..
\]

(3800)

The longest of Halliday's examples (1985:180-1) illustrates the purview of this entry: 'will have been going to have been taking'. He characterizes the tense and aspectual elements in this phrase as the combination of 'future in past in future in past in present' (1985:180-1). This example, represented in rule form, is:

\[
\text{Tns: } \text{Fut} \ VP: ([\text{Perf]} \ [Fut \ [Perf]]] \ [Prog] \ Vbl: (V.T: take NP: ?)) \ \\
\backslash \text{Time: } (\text{Fut} \ [[\text{Past}] \ [Fut \ [Past]]] \ [Pres]) \ Vbl: (V.T: take NP: ?)
\]

5.4 **Circumstantialis**

The discussion of prepositional phrases (cf.4.8, 4.9, 4.19) states that they can function as locative, temporal, manner, or probability expressions. The grammar at present accommodates the first two of these, locative and temporal. Similar rules can be stated for the two latter ones as well.
A prepositional phrase (PP) is the lexemic representation of a temporal expression (Temp) if the participant type required by the preposition X (cf. 3750-3755) is 'temporal' (cf. <temporal>) (cf. 5.1 and (3780-3782)), and the participant Y is of the required type ('temporal' designated between dollar signs). For example, this rule instantiated to 'by Christmas' is: PP: (by P: (Head: (Christmas Sg))) \ Temp: (by <temporal> P: (Head: (Christmas $temporal$ Sg)))

X can be instantiated to 'by' due to the presence of (3753) by \ end point in time <temporal>. Y can be instantiated to 'Christmas' due to the presence of (3551) Christmas \ day + etc., and (3782) day \ temporal.

The following entry has virtually the same information as the one just discussed, except that it provides for prepositional phrases which are functioning as participant attributes (Patr.PP). A separate rule is given simply for readability.

Patr.PP: (X P: (..Head: (Y X1) ..)) \ Patr.PP: (Temp: (X <temporal> P: (..Head: (Y $temporal$ X1) ..)))

The reconciliation of prepositions and their participants in locative expressions mirrors that for temporal expressions. As above, two rules are given for the sake of readability, but they have the same essential information.

PP: (X P: (..Head: (Y X1) ..)) \ Loc: (X <thing> P: (..Head: (Y $thing$ X1) ..)) (3876)
\texttt{Pattr.PP: (X P: (..Head: (Y X1) ..)) \ Pattr.PP: (Loc: (X <thing> P: (..Head: (Y $\text{thing}$ X1) ..))} \hspace{1cm} (3875)

\texttt{(3876), instantiated to 'on the car', is: PP: (on P: (the Head: (car Sg))) \ Loc: (on<thing> P: (the Head: (car $\text{thing}$ Sg)))}

\textit{X} can be instantiated to 'on' due to the presence of (3751) on\ etc. \texttt{<concrete-object>} and (3600) \texttt{concrete-object} \texttt{\ thing}. \textit{Y} can be instantiated to 'car' due to the presence of (3553) \texttt{car \ inanimate}, (3604) \texttt{inanimate \ concrete-object}, and (3600) \texttt{concrete-object \ thing}.

An adverb may also function as one of the four types of expressions discussed above, locative, temporal, manner, or probability, or as a degree or frequency expression. The following set of entries represents these facts.

\texttt{Adv.Loc: X \ Loc: X $\text{locative}$} \hspace{1cm} (3900)
\texttt{Adv.Temp: X \ Temp: X $\text{temporal}$} \hspace{1cm} (3901)
\texttt{Adv.Degr: X \ Degr: X $\text{degree}$} \hspace{1cm} (3902)
\texttt{Adv.Freq: X \ Freq: X $\text{frequency}$} \hspace{1cm} (3903)
\texttt{Adv.Prob: X \ Prob: X $\text{probability}$} \hspace{1cm} (3904)
\texttt{Adv.Manr: X \ Manr: X $\text{manner}$} \hspace{1cm} (3905)

The element \textit{X} must have the semantic property specified between dollar signs to be realized as an adverb similarly distinguished. The semotactic label (on the left side of the colon, on the right side of the rule) must be present to specify the element \textit{X} so that not all elements with the semantic property specified are realized as adverbs. For example, 'evening' has (indirectly) the semantic property 'temporal', but it is not an adverb.
A locative adverb such as 'here' (cf.1800) is the realization of a locative expression (Adv.Loc:here \ Loc:here $locative$), 'rapidly' (cf.1900), a manner expression (Adv.Manr:rapidly \ Manr:(rapidly $manner$)).

5.5 Modality

The '... intermediate degrees, between positive and negative poles, are known collectively as modality' (Halliday 1985:86). These intermediate possibilities can be expressed in one of three ways: (a) by a modal, 'Gret can see Plum Pudding on the porch', (b) by an adverb of probability, 'Gret probably sees Plum Pudding on the porch', or (c) by both, 'Gret can probably see Plum Pudding on the porch'.

The entry that reflects these facts follows.

[Mod:X] ..[Prob:Y] \ Modality:([X] [Prob:Y]) .. (3930)

Modality occurs as probability expressions (Prob) and modals (Mod). This rule, instantiated to example (c) above, is: Mod:can Tns:Pres VP:(Prob:probably Vbl:etc.) \ Modality:(can Prob:probably) Tns:Pres VP:(Vbl:(etc.))

The two dots are instantiated to Tns and VP for this example.

5.6 Clause structure

The following rules define the semotactic structure of different clause types. There are five entries corresponding to event types (cf.5.1) and voice.

A clause is a unit of structure which consists of, at the next lower level of structure, the subject noun phrase followed by the predicate (Pred). The rules discussed above
address the semantic interface for most of the constituents in this noun phrase and predicate such that the left sides of the following entries, which are at a slightly higher level of analysis than the preceding ones, appear, with the exception of the verb (and the lextactic labels Pred and Vbl), in their semantic representation (cf. P rather than NP as the first constituent in the entries, and Modality, Time, etc. in the predicate). The following rules address the interface between the verb and the event (Rel) it realizes. Since the event is the nucleus of a clause and presupposes the occurrence of certain other elements therein (cf. 5.1), a reconciliation of the event and the other elements occurs in these entries.

The first four entries deal with active voice. Participant one (P1) always bears the focus (Focus) in active clauses. The first rule provides for an event (Rel) with one participant.


X3 an Y4 (in lines 2 and 4 of the rule) are circumstantial (cf. 5.4). The intransitive verb (V.I) Y3 (in the second line of the rule) is the realization of a relation (Rel in line 3) which requires one participant (P1) whose type (X5) is specified in angle brackets by the relation and between dollar signs for the participant (cf. 5.1). For example, in 'The cat is sitting quietly', 'sit' is the
relation and 'the cat' is P1. The example represented in rule form is:

```
P:((the Head:(cat Sg)) Pred:(Time:(Pres Pres) Vbl:(V.I:sit) Manr:quietly) \ Cl:(Time:(Pres Pres) Rel:sit<thing> Focus:(P1:(the Head:(cat $thing$ Sg)))
Manr:quietly)). <Thing> comes from the entry for the sememe 'sit', and $thing$ comes indirectly from the sememe 'cat' which has the sememic property 'animal' which is a 'thing' (cf.(3605,3603,3600)). This clause is well-formed because the participant type presupposed by the relation (event) and the participant which occurs in the clause are sememically compatible.
```

The next rule applies to events with two participants as illustrated by this sentence: 'The dancer grabbed the barre at the end of her pirouette'. The reader is invited to interpret the entry in a manner similar to that explained above. <X6> (in line 4) indicates the presupposed participant type of P1 and <Y6> the presupposed participant type of P2 (cf.$X6$ and $Y6$ in P1 and P2, respectively, in lines 4 and 5).

```
```

(3960)

An abridged representation of the above example in rule form is (cf. Chapter 7 for complete instantiations of clauses):

```
P:((the dancer) Pred:((--- Vbl:(V.T::grab P:((the barre) ---) \ Cl:((--- Rel:grab(<animate><thing> Focus:(P1:((the dancer $animate$)) P2:((the barre $thing$)---)
```
These sentences illustrate the following event type: 'Their divertissements seemed awkward' and 'Max is funny'. The events in this clause type are limited to 'be' (V.Be) and the (simple) copulas (V.C) (cf.3.8). Both of these kinds of events take two participants; the second occurs as a predicate adjective at the lower level. Concord between the participants and participant types required by the event is as discussed above. The second participant type, 'quality', is specified rather than given as a variable since some of the events ('be' and V.C:become) can have other types of P2, but it is this particular type which is required by this clause construction.


The second example above, represented in rule form, is:

\[ P: (Head: (Max Sg) Pred: (Time:Pres Vbl: (V.Be:be PrAdj: funny)) \ Cl: (Time:Pres Rel: be(<thing><quality>) Focus: (P1: (Head: (Max $thing$ Sg))) P2: (funny $quality$)). Y5 is instantiated to 'thing' by virtue of rule (3004) for 'be'. X4 is instantiated to 'funny' and agrees with $quality$ by virtue of rule (3653).

The final active clause structure entry is for the events 'be', 'become', and the complex copulas (V.C.Cx) such as 'seem-like' (cf.3.8 and 4.20). These event types take two participants. These sentences exemplify this clause
type: 'Makarova is a prima ballerina', 'Plumkin became a fat cat', and 'BeBop looks like a sheep'.

P:(..Head:(X Y)...) Pred:([Neg] [Modality:X1] Time:X2 [Y2:X3]* Vbl:((V.Be:be, V.C:become, V.C.Cx:Y3) P:(..Head:(X4 Y4)...) [X5:Y5]*) \ [Neg] C1:([Modality:X1] Time:X2 Rel:(be, become, Y3) (<X6><thing>) Focus:(P1:(..Head:(X $X6$ Y)...) P2:(..Head:(X4 $thing$ Y4)...) [Y2:X3]* [X5:Y5]*) (3970)

The individual relations, be, become, or Y3 (V.C.Cx), determine the instantiation of X6 (in line 5). <thing> is also presupposed by the relation, but it is specified rather than given as a variable since these types of relations presuppose more than one P2 type, but it is this particular type that is required by this clause construction (cf.3004, 3005, for example).

The following entry provides for clauses with passive voice, as in 'The goldfish was eaten by the kitten'. Participant two bears the focus in passive clauses. P1 does not have to be present (Y4 and X5 are instantiated to $\emptyset$ in P1 (line 4)), but its type is known nonetheless because it is presupposed by the event (cf.<Y6> and $Y6$ in line 4).

P:(..Head:(X Y)...) Pred:([Neg] [Modality:X1] Time:X2 [Y2:X3]* Vbl:((Pass V.T:Y3) ~([X4:(by P:(..Head:(Y4 X5)..)) [X5:X6]*) \ [Neg] C1:([Modality:X1] Time:X2 Rel:(<Y6><X7>) P1:(..Head:(Y4 $Y6$ X5)...) Focus:(P2:(..Head:(X $X7$ Y)...) [Y2:X3]* [X5:Y6]*) [X4:(by P:(..Head:(Y4 X5)..)) IF P1:(Y4 = $\emptyset$)]) (3950)

On the expression side of the rule, the element 'passive' (Pass in line 2) is the realization of participant two (P2) focus. The element 'by' (in line 2), if introducing P1, is also a realization of P2 focus. 'By', however, can
be part of a circumstantial expression (cf.5.4) if P1 is not present (cf. If P1: (Y4 = Ø) in line 6) (cf.7.1). X4, Y2, and Y5 represent such labels as locative, temporal, and so forth. If the head of the prepositional phrase with the label X4 (in line 2) is not interpreted as P1, then the phrase is interpreted as a locative or temporal. An example of a passive clause in which the 'by' phrase is the realization of P1 is: 'The ballerina was supported around the waist by her partner.' An example in which the 'by' phrase is a circumstantial expression is: 'The picnic was eaten by the river'. P1, the participant who did the eating, is not expressed. 'By the river' is a locative expression.

The transitive verb Y3 (in line 2) is the realization of a relation (in line 4) which requires two participants, expressed by the variables Y6 and X7 in angle brackets after the relation.

5.7 **Clause types**

The grammar supports three main clause types distinguished by these labels: IntrCl for wh-questions, RelvCl for relative clauses, and DeclCl for declarative clauses. In conjunction with the different clause types are the polarity types, neutral (Ø), negative (Neg), and interrogative (Q) (polarity terms are from Halliday (Kress 1976:46)). 'Polarity is the choice between positive and negative ...' (Halliday 1985:85).
Intr [Q] [Neg] Cl:X \ [Polarity:Neg] IntrCl:X (3990)
Relv [Neg] Cl:X \ [Polarity:Neg] RelvCl:X (3991)
Ø [Q] [Neg] Cl:X \ [Polarity:[[Q] [Neg]]] DeclCl:X (3992)

(3990) takes care of wh-questions, (3991) relative clauses, and (3992) simple declaratives and yes-no questions.

Clause types expressed in entries (3990) and (3991) can have negative polarity as illustrated by these sentences:
'Whom didn't Rudolph look in on?' (Polarity:Neg IntrCl:(Rudolph Past look-in-on who) (cf.7.3 for the unabridged representation of this example in rule form)), and (the relative clause in) 'The girl who was not very big danced circles around the others' (Polarity:Neg RelvCl:(who Past be very big)).

Declarative clauses may be either neutral, negative, or interrogative, as in: 'The pianist slowed down the tempo', 'The pianist did not slow down the tempo' or 'Did the pianist slow down the tempo?' They may also have both types of marked polarity, 'Neg' and 'Q', in, for example, 'Didn't the pianist slow down the tempo?'

The entries presented in this and the last two chapters constitute the linguistic information in the present grammar. We now turn our attention to some processing considerations which use the linguistic information.
NOTES

Chapter 5

1 In his dissertation, Bennett says that the component 'surface' is not an appropriate term to use for the temporal sense of 'on', as in 'We got together on Monday.' He leaves as an open question the proper term to describe for time that which 'surface' describes for space (the other property is 'locative' whether for time or space) (cf. 1970:236, 239). I simply leave it as 'surface' for both temporal and spacial meanings.
CHAPTER 6
DYNAMIC ASPECTS

6.0 Introduction

One of the basic tenets of the view of natural language being followed in this thesis is that it is an adaptive system subject to continuous modification. An artificial system, to be functionally realistic, ought to simulate this adaptive capacity of the human linguistic system. In this chapter, a few suggestions are made for a procedure which modifies the dictionary entries of the preceding three chapters.

We distinguish in this discussion between two kinds of system modification. The first type, presented in Section 6.1, consists of adding and refining words, idioms, and constructions at the appropriate level in the grammar to permit the processing of utterances containing previously unencountered items. Lamb states that 'the process of new structure building ... goes on regularly in the lives of intellectuals ... students ... (and) ordinary adults of various walks of life... For children, ... this essentially creative process takes place much more commonly.' 'Thus there is no boundary between language acquisition and the ordinary use of language. Language acquisition is a process of structure building, and it can continue throughout life.' (Lamb, forthcoming)

The second type of system modification, discussed in
6.2, takes place at the sememic level where '...structure building occurs in the interpretation of virtually every phrase... The newly built structure is what, for the interpreter, represents the information received.' (Lamb, forthcoming)

Two major applications for an automated system are presented in 2.5: one uses the system as a rote translation device, the other as an knowledge base in which to store and from which to retrieve information. The dynamic capabilities of a system figure more importantly in an knowledge base application, although rote translation applications also require some adaptive capabilities. Those aspects of system updates pertaining to both types of applications are discussed in Section 6.1. Other facets of change more suitable to a knowledge base application are presented in 6.2.

Throughout the following discussion, we often anthropomorphize the system in describing the actions that it takes. Any "recognizing", "assessing", "remembering", and so forth attributed to the system should be understood to be the machine simulation of these human acts.

The reader is reminded that the dictionaries are used during processing, but the linguistic information in them is not procedural. When, therefore, it is sometimes stated in the discussion that processing is going on in one of the dictionaries, this is simply a paraphrase of the actual procedure (cf.2.4).
6.1 Adding and refining words and phrases

New words or new uses of words are generally encountered in the decoding process. Most humans operate on the assumption that the input sentence is acceptable, and there is a gap in their information system if they can not decode it. The system ought to operate on this supposition as well.

Adding or changing structure means adding new entries to the dictionaries or revising existing entries. These processes may be termed acquisition and refinement, respectively. If a new lexical item is encountered, then appropriate representations of it must be added to both the Morphicon and Semicon. If the current analysis of a lexical item already in the system is not compatible with some use of it in an utterance which is being decoded, then a change is made to accommodate this new information after a verification process has taken place. For example, if the system has only encountered the word 'look' as a verb, as in 'He looked in the window', and then it tries to decode an utterance like 'She gave him a look that would kill', it must refine its linguistic knowledge to accommodate this new usage. In essence, the Morphicon entry must be updated to include a tactic label of count noun, and a new Semicon entry would be made for the meaning of this word. Similarly, changes to the Lexicon are made when new constructions are encountered. For example, a construction illustrated by 'That cats are wonderful companions is well known', in which
a that-phrase serves as subject, can be added to the Lexicon when it is encountered. Accomplishing this updating of the system is no easy task, however.

The ordinary user of a natural language processing system differs from the linguist who has authored the system in that he can not readily supply the information required by the system to properly classify new items so that they will function correctly. He can not do so because the linguistic information which he possesses and uses so naturally is largely opaque to him. The system must, therefore, use the linguistic information that it already has (cf. Chapters 3, 4, and 5) to determine the classification of new items and rely on the user in only those ways in which a human interlocutor would.

The following discussion outlines a strategy for the enhancement of the system using nouns to illustrate the procedure. The exact routines are particular to nouns, but the general procedure is one applicable to all parts of speech. The usefulness of this particular procedure is far reaching since nouns are probably by far the most common addition to the linguistic system after childhood, during which time most prepositions, pronouns, determiners and the like have been acquired and refined.

Two examples, 'The guess was correct' and 'The guesses were correct' will help illustrate the points in this presentation. We refer to these examples as Input-1 and Input-2, respectively. The decoding of one input is
separate from the decoding of the other, however. For the
sake of argument, 'guess' is not presently in an entry in
the Morphicon; i.e., it is the new lexical item, information
about which we wish to place in rules in the Morphicon and
Semicon.

Since there is no entry in the Morphicon for 'guess' or
'guesses' per se, the system tries to apply rules (1980)
through (1987), which cover inflected and contracted forms.

These rules are repeated below for reference.

(1981) N.X:Y<-s,-z,-x,-ch,-sh,-y/i> -e s \ N.X:Y Nmr:Pl
(1982) V.X:Y -s \ C:3-Sg Tns:Pres V.X:Y
(1983) V.X:Y<-s,-z,-x,-ch,-sh,-y/i> -e s \nC:3-Sg Tns:Pres V.X:Y
(1984) V.X:Y -d \ (Tns:Past, en) V.X:Y
(1985) V.X:(Y, Y<-y/i>, Y<-X1/X1X1>) -e d \n(Tns:Past, en) V.X:Y IF X1 = (b,d,g,p,t,m,n,1,r)
(1986) V.X:(Y, Y<-e/∅>, Y<-X1/X1X1>) -i n g \ ing V.X:Y
IF X1 = (b,d,g,p,t,m,n,1,r)
(1987) (Aux:X, Mod:Y [Tns:X1], V.Be:be) -n ' t \n(Aux:X, Mod:Y [Tns:X1], V.Be:be) not

These rules require the presence of another rule in the
Morphicon which will instantiate the variables on the
expression sides of the rules (cf.3.1,3.8,7.1), but no such
rule is available. For example, rules (1980) and (1982)
require there to be an entry for 'gues' (these rules excise
the final -s of 'guess') as a noun and verb, respectively;
however, no entry exists for this reduced form. Rules
(1981) and (1983), when applied to 'guesses', both separate
-es from 'guess', but, although this is the desired analy-
sis, there is as yet no entry for 'guess' in the Morphicon.
At this point, the system recognizes that it has a new word,
and takes the following actions.

The procedure relaxes its requirement that rules (1980) through (1987) be accompanied by another rule which instantiates their variables. Although using these rules in this way will produce some spurious parses, all of these possibilities must be entertained until the correct one is ferretted out. The following provisional analyses are therefore produced.

For 'guess':

(1982) C:3-Sg Tns:Pres V.X:gues

For 'guesses':

(1980) N.X:guessse Nmr:Pl
(1982) C:3-Sg Tns:Pres V.X:guessse
(1983) C:3-Sg Tns:Pres V.X:guess

Since rules (1980) through (1987) deal only with nouns and verbs, some provision for the possibility that 'guess' or 'guesses' is a different part of speech must be made. There is also the possibility that these inputs are indeed nouns or verbs, but uninflected and therefore not subject to rules (1980) through (1987). To compensate, the procedure assigns a variable as a tactic category and adds each of the following to its respective output streams.

For 'guess': X:guess
For 'guesses': X:guesses

The unfamiliar input is converted by the procedure to an acceptable representation by creating a symbol for the morpheme which consists of the graphemes from the input string, and using a variable X as its tactic code. Having
done this, the procedure recommences ordinary parsing procedures until it reaches the end of the input stream. The provisional analyses from the Morphicon for each of the inputs is given below, labelled for the sake of reference as Rep-1 and Rep-2 (Rep = representation). The analyses for 'guess' and 'guesses' presented above are enclosed in parentheses and separated by commas in the following.

Input-1: The guess was correct.
Rep-1: Det.Def:the Nmr:(Sg,P1) (N.X:guess Nmr:P1, C:3-Sg Tns:Pres V.X:guess, X:guess) C:3-Sg Tns:Past V.Be:be Adj.Subj:correct

Input-2: The guesses were correct.
Rep-2: Det.Def:the Nmr:(Sg,P1) (N.X:guesses Nmr:P1, N.X:guess Nmr:P1, C:3-Sg Tns:Pres V.X:guesses, C:3-Sg Tns:Pres V.X:guess, X:guesses) C:3-Sg Tns:Past V.Be:be Adj.Subj:correct

Nothing definitive has been produced so far by the strategy for discovering the tactic category of the unfamiliar item. A much more reliable clue to the classification of a new word comes from its context in the string. Humans learn to use new words based on their lextactic distribution, so this is a method the system can use as well.

To accomplish these ends, the Lexicon is entered and all possible rules are tried. Using entry (2100), stated below, produces the following provisional analyses. The letters in parentheses at the right margin are for reference.

(2100) N.X:Y Nmr:(P1,Ø) \ N.X:Y Nmr:(P1,Sg)

Rep-1: N.X:guess Nmr:P1 \ N.X:guess Nmr:P1 (a)
X:guess \ N.X:guess Nmr:Sg (b)
Rep-2:  \(N.X:\text{guesse Nmr:P1} \setminus N.X:\text{guesse Nmr:P1}\) (a)
\(N.X:\text{guess Nmr:P1} \setminus N.X:\text{guess Nmr:P1}\) (b)
\(X:\text{guesses} \setminus N.X:\text{guesses Nmr:Sg}\) (c)

Rule (2110) applies to the possibilities that 'guess' is a verb. We do not describe any paths from this point on, however, that do not pertain to nouns, but the reader should remain aware that the system is intended to entertain all possibilities until the solution is found. The blocking of spurious paths is discussed in 2.4.

Rule (2125) (cf. Appendix A) applies to all of the abovementioned provisional analyses, forming nominals. Finally, the noun phrase rule (cf. 2131 reproduced in part below) with a definite determiner followed by a nominal is encountered, and the following possibilities are produced. The letters (a), (b), and (c), at the right below, correspond to the same letters in the representations given above.

\[(2131) \quad \text{Det.Def:} X \text{Nmr:Y Nom: (Head: (N.X_1:Y_1 Nmr:Y))}\]
\(\setminus \text{NP: (X Head: (Y_1 Nmr:Y))}\)

Rep-1:
\[
\text{Det.Def: the Nmr: (Sg,P1) Nom: (Head: (N.X: guesse Nmr:P1))}\]
\(\setminus \text{NP: (the Head: (guesse Nmr:P1))}\) (a)

\[
\text{Det.Def: the Nmr: (Sg,P1) Nom: (Head: (N.X: guess Nmr:Sg))}\]
\(\setminus \text{NP: (the Head: (guess Nmr:Sg))}\) (b)

Rep-2:
\[
\text{Det.Def: the Nmr: (Sg,P1) Nom: (Head: (N.X: guesse Nrm:P1))}\]
\(\setminus \text{NP: (the Head: (guesse Nrm:P1))}\) (a)

\[
\text{Det.Def: the Nmr: (Sg,P1) Nom: (Head: (N.X: guess Nmr:P1))}\]
\(\setminus \text{NP: (the Head: (guess Nmr:P1))}\) (b)

\[
\text{Det.Def: the Nmr: (Sg,P1) Nom: (Head: (N.X: guesses Nmr:Sg))}\]
\(\setminus \text{NP: (the Head: (guesses Nmr:Sg))}\) (c)

The noun in rule (2131) does require a sub-type (cf. X1). Had this rule specified a particular sub-type rather
than representing it as a variable (which indicates that any sub-type of noun is permissible in this construction), that particular sub-type would have been assigned at this time to 'guess'.

To recapitulate briefly, the two inputs 'guess' and 'guesses' have been assigned the tactic code noun of unknown sub-type based on lexotactic co-occurrence restrictions (i.e., the immediate constituents matched the noun phrase rule but no other rule). By virtue of such restrictions, the possibility of the inputs' being assigned to other tactic categories such as verb (mentioned above and discussed in 2.4) has been eliminated.

At this point, the procedure has gotten about as far as it can on its own. It therefore enters an interactive mode and holds a consultation with the user.

One aim of this consultation is to determine which of the entries (a) or (b) of Rep-1 correctly represents 'guess' (or (a), (b), or (c) of Rep-2 for 'guesses'). This can be done by finding out if the noun is singular or plural since that is one of the differences between (a) and (b) of Rep-1. The reader may wonder why the procedure simply did not ask the user right away for this (plurality) information. The reason is that such a question was not appropriate until the procedure had determined that the input was a noun, and this was not accomplished until the noun phrase rule matched the input, and the other possibilities were blocked out (because they did not match a phrase structure entry).
Another aim is to determine the noun sub-type. The procedure may accomplish both tasks by having the user respond to some questions. The following dialogue illustrates the interchange between system (S:) and user (U:).

For Input-1:

\[ S: \text{You said, "The guess was correct."} \]
\[ \text{I don't know the word 'guess'. Will you please answer some questions to help me understand 'guess'?} \]
\[ \text{(Type Y,N)} \]

\[ U: Y \]

\[ S: \text{Type in the number of every sentence which looks to you like the way you would ask a question in ordinary English. If none of these looks okay, type in 'Ø'.} \]

\[ \text{Can I say:} \]
\[ (1) \text{What is a guess?} \]
\[ (2) \text{What is guess?} \]
\[ (3) \text{What are guess?} \]

\[ U: 1 \]
For Input-2:

S: You said, "The guesses were correct."
I don't know the word 'guesses'. Will you please
answer some questions to help me understand
'guesses'? (Type Y,N)

U: Y

S: Type in the number of every sentence which looks to
you like the way you would ask a question in
ordinary English. If none of these looks okay,
type in '0'.

Can I say:
(1) What is a guesses?
(2) What is guesses?
(3) What are guesses?

U: 3

S: Can I say:
(4) What is a guesse?
(5) What is a guess?

U: 5

A positive response to choice one (1), 'What is a
___?', indicates that the input noun is count and singular.
Choice two (2), 'What is ___?', signifies a mass noun. Mass
nouns are always considered singular (cf.3.1). Choice three
(3), 'What are ___?', designates a plural count noun. Pro-
per nouns and hour nouns are not considered for 'guess' or
'guesses' due to the orthographic conventions associated
with proper names (cf.2.4 and 3.1) and the limited domain of
that which is acceptable as an hour noun (cf.4.20).

The system poses all the possibilities it recognizes as
having potential to the user (cf. (1), (2), and (3), above)
rather than simply presenting these sample questions serial-
ly and quitting after it gets a positive response to one.
This practice allows the user to pick out for the system all possibilities for noun sub-types. The reader will recall that some nouns belong to more than one sub-category (cf. 'time' (1081) in 3.1). Of course, the user may not pick out every possibility even if more than one exists, since he probably has the one which fits with the usage in the input sentence in mind. But this is all right. The entry in the Morphicon can be amended later if the need arises.

In the interaction between system and user for Input-1 above, the user identifies for the system that 'guess' is a singular count noun (cf. user choice 1). The system therefore recognizes Rep-1 (b) above as the desired path, and it stops processing Rep-1 (a). Having thus acquired all of the necessary information to make an entry to the Morphicon, the following rule is added.

```
g u e s s \ N.Cnt:guess
```

This Morphicon entry does not specify singular number for 'guess' even though we know that that is how it was used in Input-1. All nouns not specifically marked for plural in the Morphicon are, by default, marked as singular in the Lexicon (cf.4.2).

In the system/user dialogue for Input-2 above, 'guesses' is found to be a plural count noun (cf. user choice 3). This eliminates Rep-2 (a) above, leaving 'guesse' or 'guess' as the singular form of 'guesses' (cf. Rep-2 (b) and (c), respectively). The user is presented with a choice between these possibilities and chooses
'guess' (cf. user choice 5). The system makes this entry in the Morphicon.

    guess s \ N.Cnt:guess

Rule (1981) for -es suffixes on nouns supplies the plural number for the use of this noun in Input-2.

Had the plural form of the noun been a so-called irregular form such as 'wives', 'men', 'children', etc., the attempt by the system to pinpoint the singular form would have been futile. Thus, if the system gets a 0 response to its query where it got a 5 above, it elicits the singular form from the user and places that, along with plural number, in the entry it adds to the Morphicon (cf. 3.1 - the treatment of irregular inflected forms). A second entry for the singular form would also be added.

There are other clues to the number of a noun, such as subject-verb agreement (cf.4.13) or predicate nominative-antecedent agreement (cf.4.8), etc., and the system could use these rules if the system author desires. Since these clues are not available in every input, however, we rely on the procedure outlined above.

After making the Morphicon entry, the next step in the procedure of processing an unfamiliar item is to make a Semicron entry so that decoding may continue and understanding (cf.6.2) may result.

In 5.1, it is argued that nouns are names for 'things', and those 'things', which are actually concepts for 'things', could be sub-categorized and located in a
taxonomic hierarchy according to their direct and indirect relations to other concepts. Participation of concepts in semotactic constructions is restricted by their membership in the classes in such a hierarchy (cf. 5.2, 5.4, 5.6).

In addition to being categorized as a member of a general class such as concrete-object or abstract-object, the concept must be differentiated from other members of that class. One way of achieving the former goal is to ask the user for a synonym. A dialogue like the following might ensue.

S: Can you substitute another word for 'guess' in the sentence, "The guess was correct"? Please give me some examples. (Type Ø for no substitute word.)

U: judgement, estimate, supposition, speculation, surmise

One would hope that there is already one of these choices in the Semicon. If there is, then the meaning of that concept may be copied exactly for 'guess', and an entry for 'guess' added to the Semicon. That we do not accomplish the second goal mentioned above, that of differentiating a concept from other members of its class, in this case differentiating 'guess' from 'judgement', for example, is all right. It often takes humans years of exposure to words which are ostensibly synonymous to refine their systems such that the connections which delimit each concept have some difference between them.

If none of the synonyms given is already in the system,
or if no synonym is given, then the task of making a Semicon entry becomes more difficult. So that at least the semotactic co-occurrence restrictions can be satisfied, the procedure tries to place the concept in one of the general categories of the taxonomy.

| S: Is a 'guess' something you can physically touch? |
| U: N |
| S: Is it an emotion, like love or hate? |
| U: N |
| S: Is it a time, like Monday or the morning? |
| U: N |
| S: Is it a sensory or cognitive "object", like an idea, knowledge, or a dream? |
| U: Y |

The first user response in this dialogue tells us that a 'guess' is an abstract, rather than concrete, object. The responses to the subsequent questions (each of which tries a sub-category of abstract-objects) places 'guess' in the hierarchy as a sensory/cognitive object.

The next step is to make some connections to other concepts in the network so that some more specific meaning for this concept 'guess' can be established. Because the "area" of the network in which 'guess' is located has been delimited by the fact that it is a sensory or cognitive object, the search for some meaning does not have to go outside of that which is entailed by the other concepts of this type.
S: Is a 'guess' something you can smell?
U: N
S: Is it something you can hear?
U: N
S: Is it something you can taste?
U: N
S: Is it something you can think?
U: Y

Prompted by the positive response to the last question, the procedure recognizes that 'guess' is a 'cognitive-object'. It takes the concepts associated with 'think' and applies them to 'guess', adjusting for the fact that 'think' is an event sememe and 'guess' is a 'thing' (this information is represented as 'etc.' below). A Semicon entry for 'guess' is then made.

guess \ cognitive-object + etc.

To determine if such a strategy is viable, we consult just an ordinary dictionary. Webster's New World Dictionary gives as one of its definitions for the noun 'guess': 'a judgement or estimate formed by guessing', and 'guessing' is of course the present participle form of the verb 'guess' for which one of their definitions is: 'to think or suppose' (1980:621). Taking the opposite tack, we look at the entry for the noun 'thinking', the nominal equivalent (since 'guess' is a noun) of the verb 'think': 'the action of one who thinks'. Following this trail, we find this
definition for 'think': 'to have an opinion, judgement, etc.' (1980:1478). The decision of the system to interpret 'guess' as a kind of 'thinking' is not a farfetched one then, as indicated by the mutual references in these dictionary definitions.

With the addition of the Morphicon and Semicon entries described above, the procedure is able to complete its decoding of the input sentence(s) and thus in effect to "remember" the new information for future use. The next step in the process of understanding the decoded input is discussed in Section 6.2.

Not only do new entries have to be added, but existing entries have to be changed to keep the system current. Some suggestions on how to accomplish such refinements are therefore in order.

The input sentences to illustrate this discussion are 'He guessed it' and 'He guesses it', referred to as Input-3 and Input-4, respectively. The decoding of each of these inputs is separate from the decoding of the other.

The system has entries in the Morphicon for all of the elements in these inputs since 'guess' has now been added (cf. above discussion). The only encounter the system has had with 'guess', however, has been in its use as a noun, and that is what the Morphicon entry reflects. For Input-4, this presents no immediate problem, and the following parse, though erroneous, results.
Input-4: He guesses it.
Rep-4: Pro:3-male Nmr:Sg N.Cnt:guess Nmr:Pl Pro:3-neuter Nmr:Sg m

With Input-3, however, the application of rule (1985) requires that a verb, not a count noun, occur with the suffix -ed. The procedure takes the same two actions as previously described. One action relaxes the requirement that the variable on the left side of rule (1985) be instantiated by another Morphion rule. The procedure simply puts out a string with the input assigned to the tactic classification in rule (1985). Just the part of the input in question is given below.

Input-3: guessed
Rep-3: Tns:Past V.X:guess

(a)

The second action of the system is to assign a variable as a tactic code for the input string in the event that 'guessed' is not an inflected verb (as rule (1985) would indicate) but rather an uninflected verb or some other part of speech. The morpheme symbol is based on the graphemic input. This representation is produced.

Rep-3: X:guessed

(b)

An action similar to the second one just shown (cf.(b)) is not taken for Input-4. The system assumes, although erroneously, that it has given the correct parse since all usual requirements have been met (i.e., since rule (1981) is accompanied by a compatible entry in the Morphicon).

Rep-3 (a) is processed in the Lexicon without incident (since this is the correct parse). When it reaches the
rules for the formation of verbals, the previously unknown verb sub-type is instantiated to transitive (T) because, given the parse of the rest of the input string, conditions for rule (2201) are met, and the verb sub-type specified in that rule is transitive (V.T:guess). An abridged version of rule (2201) follows.

(2201) V.T:X NP:Y \ Vbl:(V.T:X NP:Y)

One of the paths that Rep-3 (b) has taken has allowed the variable to be instantiated to a verb of unknown sub-type as well (cf.(2110): X:guessed \ Tns:Pres V.X:guessed). This parse also fits rule (2201), and the variable is instantiated to sub-type transitive (V.T:guessed).

The other paths taken for Rep-3 (b) will have come to naught in the Lexicon, leaving for Input-3 only these two parsings which differ in their tense assignment and assessment of the root form of the verb. The system consults the user to make a decision about which interpretation is correct.

+---------------------------------------------------------------+
| S: In "He guessed it", is 'guessed' expressing a present | or past action? (Type present, past, or Ø for neither) |
| U: Past |
+---------------------------------------------------------------+

With this response, the system recognizes Rep-3 (a) as the correct interpretation, ceases processing of (b), and amends the Morphicon entry for 'guess' to look like the following.

g u e s s \ (N.Cnt, V.T):guess
There is a bigger problem with Rep-4, however. No construction in the Lexicon accommodates three consecutive noun phrases. Consequently, the system, working under the assumption that the input is a clause, goes to the rules in the Semicon and finds that the minimum requirements for a clause are a noun phrase followed by a predicate (cf. 5.6). Using this information, it reenters the Lexicon and tries to build a predicate out of the last two noun phrases. Trying to get a match on some rule, the procedure replaces first one then the other of the tactic codes of the noun phrases with a variable. It ignores number when trying tactic codes other than noun. We do not present the entire logic of this endeavor step by step, but do assert that such a strategy will ultimately lead the system to reassess 'guesses' as a verb. Rule (1983), which takes care of the verbal suffix -es, reinforces the viability of such an assessment.

Before making the Morphicon amendment, the system can verify its conclusions by consulting the user and offering some other sentences in which 'guess' is used as a verb. If they are acceptable, then the entry in the Morphicon is updated as shown above in the discussion of Rep-3.

A Semicon entry must also be made as a result of processing either of these inputs. Part of the intension of an event sememe is the specification of its participant requirements. To determine the number and type of participants which the event sememe 'guess' requires, the system looks at the input sentence participants for hints. 'He'
and 'it' have (indirectly (cf.3811)) the properties 'male' and 'thing', respectively. The object in assigning participant requirements is to cover all possible participant types without taking in unwanted categories. To do this, the system goes to the next higher category in the taxonomy of 'things'(cf.5.1). 'Humans' as well as other 'animals' dominate 'male' in the hierarchy. The system queries the user.

```
S: Can a human guess? (Type Y, N, M(aybe))
U: Y
S: Can an animal guess?
U: M
```

The equivocal answer of the user causes a break in this line of questions. The procedure will include in an entry only unequivocal information, so it instantiates the first participant required by 'guess' to 'human'. The basis for such an action rests on the assumption that only that which is necessary for the resolution of the input sentence need be included at this time, and that the user will at least give an unequivocal response to the choice that reflects the use in the input sentence.

The second participant in the input, 'it', has the property 'thing'. Since 'thing' is at the top of the taxonomy, there is no more general category to which a participant can belong. The second participant requirement of 'guess' is therefore instantiated to 'thing'.

In addition to the presupposed participants of 'guess',
the rest of the intension must be specified. This can be accomplished in a manner similar to that discussed for the 'thing' sememe 'guess'. Following a taxonomic tree of event types, the system poses questions to the user to establish, at the least, that 'guess' is a hyponym of 'do' rather than a hyponym of 'be'. We leave the exact method unstated, and use 'etc.' to indicate the findings of such a procedure.

The Semicon entry for the event 'guess' is now ready to be made. The reader will notice that the sememe for the 'thing' 'guess', already in the Semicon, is not amended to cover the event sememe; rather, a new entry is made. This is due to the fact that an event sememe and a 'thing' sememe are different, even if related, concepts. The overlap in their meanings (meaning is represented below simply as 'etc.') shows the extent to which they are related.

guess \ etc. (<human><thing>)

We now turn our attention to the problem of adding new constructions to the Lexicon and Semicon. The input sentence that is used to illustrate this procedure is 'The bee is in the box', identified hereafter as Input-5.

The procedure gives an initial parse of the input using the rules in the Morphicon.

Input-5: The bee is in the box.
Rep-5: Det.Def:the Nmr:Sg N.Cnt:bee C:3-Sg Tns:Pres V.Be:be Prep:in Det.Def:the Nmr:Sg N.Cnt:box

Entering the Lexicon, the system is able to make noun phrases of 'the bee' and 'the box' using rule (2131) (cf. above), and a prepositional phrase of 'in the box' using
Rule (2180) (abridged version: Prep:X NP:Y \ PP:(X NP:Y)). Rule (2166) (cf. Appendix A) takes care of subject-verb agreement. The two rules for verbals that specify the verb 'be' (cf. (2203) and (2204)) both require a constituent not present in this decoding, however, (2203) a predicate adjective, (2204) a predicate nominative. Ordinary processing therefore stops.

Using the same strategy discussed above for Input-4 (which was decoded initially to three noun phrases), the system, working under the assumption that what it has as input is both viable and a clause, enters the Semicon rules and discovers that the minimum requirements for a clause are a noun phrase and a following predicate. Identifying that it has in the decoding a noun phrase followed by the problematic portion of the parse, the system works under the assumption that this portion that it can not match to an existing rule is the predicate. It therefore looks at the rule for predicates (2400) and, reversing its strategy, tries to use this rule to encode.

The right side of rule (2400) has this form:
The portion of the decoding that is being matched to this rule is: Tns:Pres V.Be:be PP:(in the box). Y is therefore instantiated to 'present' (Pres) and Y2 to 'in the box', which leaves as the only possible assignment of 'be' X1, a verb phrase. 'Be' can not be X2 because X2 is optional in this rule while X1 is not.
The system goes next to the verb phrase rule (2300), and identifies V.Be:be as a verbal. It then goes to the rules for verbals and find two for 'be', the same two it encountered earlier in the decoding of Input-5 that caused ordinary processing to stop.

Having arrived at the same point in the rules using both decoding and encoding approaches, the system guesses, incorrectly, that a verbal may consist of the verb 'be' alone and makes a Lexicon entry to that effect.

V.Be:be \ Vbl:(V.Be:be)

For every verbal rule in the Lexicon, there is a corresponding clause structure rule in the Semicon. The system therefore adds a rule to the Semicon to accommodate this new verbal, modelling it after the two rules for 'be' that are already there (cf.(3965) and (3970)) but adjusting for the elements it has decoded in Input-5.

P:(.Head:(X Y)...) Pred:([Neg] [Modality:X1] Time:Y1 Vbl:(V.Be:be) [Loc:X2]*) \ [Neg] Cl:([Modality:X1] Time:Y1 Rel:be<thing> Focus:(P1:(.Head:(X $thing$ Y)...) [Loc:X2]*)

(The verbal (Vbl) and relation (Rel) are different from (3965) and (3970), and there is no participant two (P2)).

Future exposure to other inputs will prompt the system to include in the verbal as a non-optional constituent the locative prepositional phrase (or locative adverb) (cf.X2) that follows 'be'. When it does, the Lexicon and Semicon entries just added will be updated to reflect that information, and the Semicon entry for the sememe 'be' will be
amended to include 'locative' as another kind of second participant.

be \ etc. (<thing> ((thing,quality,locative)>)

That this is not done immediately is realistic, however, in the sense that humans too must have multiple exposures to certain phenomena to make a generalization.

The preceding discussion as well as the one which follows are, of course, just a few suggestions and not an attempt at a complete survey of strategies to modify the system.

6.2 Adding, verifying, and retrieving information

The type of system modification discussed in this section is meant to simulate human understanding. After some input has been decoded to the clause level, further actions must still be taken or the system, as sophisticated as it might be, is rather useless. Humans use their systems to store, verify, and retrieve information, and the model of the system ought to perform in a similar manner.

The first task of the system is to build a high level representation of the decoded clause. The procedure connects the sememes within the clause to the more general concepts of which they are particular instances (cf.1.2). It also maintains the integrity of the semotactic relationship between the sememes in the clause. For example, in 'The cat ate a bird', 'cat', 'eat', and 'bird' are in a particular (semotactic) relationship to one another; that is; they form a proposition. They are also specific instan-
ces of the general (class) sememes 'cat', 'eat', and 'bird', respectively. We return to the details of conceptual structure building below.

The second task of the system is to compare the newly built structure to other such structures, previously built and stored by the system. For example, after decoding 'The cat ate a bird' and building its conceptual structure, the system compares this representation to those in its existing knowledge base.

The first two actions, building and comparing, are done for every decoded clause. The next action that the system takes depends on the type of clause that was decoded. If the decoded clause is a declarative whose polarity is "neutral" (cf. 5.7), the system adds the structure to its memory (cf. 2.3). If the decoded clause is a declarative with "interrogative" polarity (cf. 5.7) (i.e., a yes-no question), the desired response is the return of an answer, either yes or no, to the user. If the decoded clause is a wh-question, then the questioned element is retrieved from the existing information in the system and returned to the user.

Some parenthetical remarks about sememes will help illustrate the following discussion. In 1.2, a figure similar to the one below is presented to explain the possibly confusing uses of the term sememe.
A sememe like 'cat' is the point of convergence of many other sememes such as feline AND furry AND pet AND so forth. 'Cat' is the sememe that represents the general class and is referred to directly only in a generic reference such as 'The cat is a beautiful creature' (cf. 3.2 and 4.6). Ordinarily, the lexical item 'cat' is used to denote a particular cat, a specific reference, as expressed in 'My cat has big ears' (cf. 3.2 and 4.5) rather than the general class sememe 'cat'. To signal that difference, the specific reference is represented as 'cat-1'.

As mentioned above, the first task of the system is to build structure. The high level representation made for all decoded clauses connects the particular sememes in the decoding to the class sememes of which they are instances. This connection is represented by the use of the same symbol (such as 'cat') for both the class and the particular instance. To differentiate one particular instance from another, a unique number is suffixed to each, for example, 'cat-1', 'cat-2', ... 'cat-306', etc. Not only 'thing' sememes are treated in this manner, but event and other sememes are as well. This representation not only provides
the connection between the sememes in the clause and their meanings (via the class sememe), but also makes it possible for the system to specify the semotactic relations among the sememes in the decoded clause. The system simply keeps a record of these uniquely identified sememes and their relations. Such a representation has been termed "gist" (Sandford and Garrod 1981:67-69).

Not every sememe is worth marking in this way since some, like time and determiners for example, are controlled by the speech act itself. Exactly how these elements are to be handled is not included in this discussion, however (cf. Lamb, forthcoming, for suggestions).

The second task of the system is to compare this newly built structure representing the information received (Lamb, forthcoming) to the existing structures which were previously built and stored by the system. Comparison is simply a search and match procedure in the knowledge base.

For decoded clauses which are declarative with "interrogative" polarity (yes-no questions), the result of the comparison procedure will determine the answer to be returned to the user. If the search procedure makes a match, that is, the decoded information matches some stored information, the answer returned is "yes". For example, if the user asks the system 'Does Plum Pudding like to swim?', and a representation of 'Plum Pudding likes to swim' is stored, then the answer to the question is "yes". If the comparison procedure finds a representation whose only difference is
that it is negative (i.e., has negative polarity (cf.5.7)), then the system returns the answer "no". For example, the representation of 'Plum Pudding doesn't like to swim', found during the search, would prompt a "no" response to the above query. If the comparison turns up no like representation, then the system returns the answer "I don't know".

Decoded clauses which are declarative are added by the system if the comparison procedure finds no pre-existing or conflicting (for example, a negative clause) representations. Should it find a conflicting statement, a routine to encode that representation, preface it with something like "I thought that", and return it to the user is in order. For example, if the system decodes 'Plum eats lizards' and the comparison procedure finds a representation of 'Plum eats bugs', then the system returns: "I thought that Plum eats bugs". The user may then give the system the further information it needs to update its information store properly (either Plum eats both; Plum eats only lizards; Plum used to eat bugs but now eats lizards, etc.).

When a wh-question is decoded and a match is made by the comparison procedure, the asked-about portion of the clause representation (i.e., the interrogative element such as 'who', 'how', 'why', etc.) is retrieved from the "found" clause and returned to the user. For example, if the user asks the system 'Who ate the pie?', and the system has stored 'Plum Pudding ate the pie', then 'Plum Pudding' is retrieved and returned to the user. Of course, if the
search procedure does not turn up a like representation, then "I don't know" is returned to the user.

If more than one decoding is produced, the system ought to have a routine to query the user to resolve the ambiguity since more than one decoding is not desirable (except in the case of puns). Such a routine would encode the two (or more) possibilities and return them to the user, who could then choose the one which reflects the meaning he intended.

These conceptual representations are stored in the Semicon, along with all of the other sememic information. They are, of course, at the higher of the two sub-levels in the dictionary (cf. 2.1). Clause structure rules constitute the lower level, meaning and stored propositions the higher.
NOTES

Chapter 6

1  "Word" is used in this discussion in its non-technical sense to mean a lexical item.

2  The majority of the ideas for this discussion are based on Lamb, 'On Constructing the Context of a Text', forthcoming, and on lectures given by him at Rice Univ. from 1983-present.

3  This information may be recorded in whatever manner that is internally efficient to the system (numbers, strings, pointers, etc.)

4  Lamb provides for a Gnosticon which is at a higher level than the Semicon to house information representations. (cf. Chapter 2 fn.1).
CHAPTER 7
ENCODING AND DECODING

7.0 Introduction

In this chapter, we step through the processes of encoding and decoding using the linguistic information in the three dictionaries elaborated in Chapters 3 - 5. One example conceptual representation of a decoded clause is given using the method outlined in 6.2 (cf. 7.1, parse 2), but the concentration is otherwise on the decoding and encoding processes to and from the clause level.

"Parse" is the term used in computation to mean 'recognizing sentences and assigning them structures' (Winograd 1983:88). The terms parse and decoding are used interchangeably in the following section to mean the interpretation of sentences to the clause level. Understanding, as defined in 6.2, is therefore excluded from the scope of the term parse.

As explained in 2.4, processing is done in a parallel fashion. All possibilities are considered, and any rule that can apply, will apply. All inappropriate paths taken during processing are winnowed out along the way. If more than one decoding is produced, this signifies true ambiguity and requires further information for its resolution (cf. 6.2).

With parallel processing (or backtracking), traces balloon significantly and soon become unwieldy. As a
consequence, attempting to follow a complete trace of a processing is often difficult. For this reason, the traces of the parses and generations presented in this chapter show only the correct path. Since only the correct paths are presented, any differences between decodings and encodings are eliminated. All of the traces are therefore shown as decodings, and the reader is invited to read them in reverse order for encoding.

Diagrams of the decodings are given after the traces of the first two examples to help the reader initially in deciphering the traces.

All of the rules used in encoding and decoding have been explained elsewhere and are therefore not further elaborated in this chapter. In a trace, the number of the rule which is being applied is stated, as is the section in Chapters 3, 4, or 5 in which it is explained. In addition, each application of a rule in a trace is numbered for reference, beginning with one (1) for each new trace.

The reader is reminded that the rules in the dictionaries apply in any order, not just in the order they are presented in in Chapters 3, 4, and 5, or in the numerical order indicated by their 4-digit reference code. The rules apply when the current input string, during any phase of the processing, matches that which is indicated in the rule (cf. 2.4).

The text of this chapter is divided into three sections: one on statements and two on questions. All to-
gether, the sentences of these clause types include a sampling of nearly all of the constructions discussed in the preceding chapters.

7.1 **Statements**

The first sentence to be parsed is a simple declarative with neutral (unmarked) polarity. The trace shows agreement between the subject and verb (line 10 of the parse), and the reconciliation of lexotactic labels and sememic properties at the semantic interface (lines 15 and 18).

For decoding, the system is entered through the Morphicon. Each word is looked up in the Morphicon by matching the input to the left side of an entry (cf.2.4). When a match is made, the right side of that entry is the output, creating a string of elements with tactic labels.

**DECODE:** The car is black.

**ENTER MORPHICON:**

1. (1200) t h e \ Det.Def:the Nmr:Sg
2. (1028) c a r \ N.Cnt:car
3. (1403) i s \ C:3-Sg Tns:Pres V.Be:be
4. (1761) b l a c k \ Adj.Col:black

The output string from the Morphicon is:

Det.Def:the Nmr:Sg N.Cnt:car C:3-Sg Tns:Pres V.Be:be Adj.Col:black

**ENTER LEXICON:**

5. (2100) assigns number to noun:
6. (2110) assigns tense to verb: (fn.1)
7. (2120) recognizes adjective phrase:
8. (2120) recognizes adjective phrase:

(3.2) (3.1) (3.8) (3.5) (4.2) (4.10) (4.3)
(2125) recognizes nominal expression:  \( \text{N.Cnt:car Nmr:Sg \ Nom: (Head: (N.Cnt:car Nmr:Sg))} \)

(2131) recognizes specific noun phrase with 'the':  \( \text{Det.Def:the Nmr:Sg Nom: (Head: (N.Cnt:car Nmr:Sg))} \ \text{NP: (the Head: (car Nmr:Sg))} \)

(2166) recognizes subject verb agreement:  \( \text{NP: (the Head: (car Nmr:Sg)) C:3-Sg Tns:Pres \ NP: (the Head: (car Nmr:Sg)) Tns:Pres} \)

(2203) recognizes verbal with 'be' and predicate adjective:  \( \text{V.Be:be AP: (Adj.Col:black) \ Vbl: (V.Be:be AP: (Adj.Col: black))} \)

(2300) recognizes verb phrase:  \( \text{Vbl: (V.Be:be AP: (Adj.Col:black)) \ VP: (Vbl: (V.Be:be AP: (Adj.Col:black)))} \)

(2400) recognizes predicate:  \( \text{Tns:Pres VP: (Vbl: (V.Be:be AP: (Adj.Col:black))) \ Pred:} \ \text{Tns:Pres VP: (Vbl: (V.Be:be AP: (Adj.Col:black)))} \)

The output string from the Lexicon is:  \( \text{NP: (the Head: (car Nmr:Sg)) Pred: (Tns:Pres VP: (Vbl: (V.Be:be AP: (Adj.Col:black))))} \)

ENTER SEMICON:

(3800) recognizes time:  \( \text{Tns:Pres VP: (Vbl: (V.Be:be AP: (Adj.Col:black))) \ Tns:Pres Vbl: (V.Be:be AP: (Adj.Col:black))} \)

(3825) recognizes semantic interface of adjective: \( \text{AP: (Adj.Col:black) \ PrAdj: (black $color$)} \)

(3651) recognizes semantic property of sememe:
black \ color

(3830) recognizes semantic interface of adjective phrase:
(fn.1) \( \text{PrAdj: black \ PrAdj: black} \)

(3925) recognizes noun phrase as participant:  \( \text{NP: (the Head: (car Nmr:Sg)) \ P: (the Head: (car Sg))} \)

(18) (fn.2) \( \text{PrAdj: black \ PrAdj: black} \)

(3965) recognizes active clause with 'be' + predicate adjective:
\( \text{P: (the Head: (car Sg)) Pred: (Time:Pres Vbl: (V.Be:be PrAdj: black)) \ Cl: (Time:Pres Rel: be (<thing> <quality>)) \ Focus: (P1: (the Head: (car $thing$ Sg)) P2: (black $quality$))} \)
(3651) black \ quality
(3553) car \ concrete-object
(3600) concrete-object \ thing

(3990) recognizes declarative clause:

(19) Cl:(Time:Pres Rel:be (<thing><quality>) Focus:(P1:(the Head:(car $thing$ Sg))) P2:(black $quality$)) \ DeclCl:(Time:Pres Rel:be Focus:(P1:(the Head:(car Sg))) P2:black)

All of the co-occurrence restrictions have been satisfied (cf. the rules that have applied and fn.2) and a clause built. This is the point in the processing in which the procedure described in 6.2 takes over and makes a conceptual representation of the decoded clause. As explained in 7.0, however, this chapter is primarily concerned with decoding to the clause level, so we do not show the conceptual representation.

The following diagram shows the trace just given in a different format. The numbers at the left side of each line correspond to the reference numbers in the trace. The scope of each rule is indicated by the dotted line.

Decode: The car is black.

<table>
<thead>
<tr>
<th>1-4</th>
<th>t h e------</th>
<th>c a r------</th>
<th>i s--------</th>
<th>b l a c k------</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-7</td>
<td>Def.Det,Nmr</td>
<td>N.Cnt-----</td>
<td>C,Tns,V.Be--</td>
<td>Adj.Col------</td>
</tr>
<tr>
<td>8</td>
<td>N.Cnt,Nmr</td>
<td>------------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
<tr>
<td>9</td>
<td>Nom--------</td>
<td>------------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
<tr>
<td>10-11</td>
<td>----------NP----------</td>
<td>------------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
<tr>
<td>12</td>
<td>----------S-V agr----------</td>
<td>------------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
<tr>
<td>13</td>
<td>----------VP----------</td>
<td>------------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
<tr>
<td>14</td>
<td>-----------Pred-----------</td>
<td>------------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
<tr>
<td>15-16</td>
<td>Time-----------Vbl--------</td>
<td>------------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
<tr>
<td>17</td>
<td>----------P----------</td>
<td>------------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
<tr>
<td>18</td>
<td>----------Cl----------</td>
<td>------------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
<tr>
<td>19</td>
<td>----------DeclCl----------</td>
<td>------------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
</tbody>
</table>

In the following traces, none of the Semicon entries
which provide the sememic properties that are required for
semantic relations are given as they were for the trace
above (cf. (15), (18) (3651, 3553, 3600)). The reader may
assume their presence since the entries in the traces have
been instantiated. For example, since in the application of
a rule for clause structure (cf. 3650-3670) the participant
types are instantiated for the relation, then that relation
has an entry in the Semicon which specifies those presup-
posed participants (cf. 5.1).

The next sentence parsed is also a declarative with
neutral (unmarked) polarity. It is more complex than the
last example, however. It has two noun phrases of specific
reference, a simple tense, and a participial phrase (an
embedded clause) that modifies one of the noun phrases.

DECODE: The two cats smelled lunch being prepared.

ENTER MORPHICON:

(1) (1200) the \ Det.Def:the Nmr:Sg (3.2)
(2) (1602) two \ Num.Card:two Nmr:Sg    (3.4)
(3) (1980, 1020) cat \ N.Cnt:cat Nmr:Pl (fn.2) (3.1)
(4) (1985, 1449) smelled \ Tns:Past V.T:smell(fn2) (3.8)
(5) (1080) lunch \ N.Mass:lunch           (3.1)
(6) (1401) ing \ V.Be:be                  (3.8)
(7) (1985, 1470) prepared \ V.T:prepare (fn.2) (3.8)

The output string from the Morphicon is:
Tns:Past V.T:smell N.Mass:lunch ing V.Be:be en V.T:prepare

ENTER LEXICON:

(2100) assigns number to noun: (4.2)
(8) N.Cnt:cat Nmr:Pl \ N.Cnt:cat Nmr:Pl (fn.1)
(9) N.Mass:lunch \ N.Mass:lunch Nmr:Sg

(2110) assigns tense to verb: (fn.1) (4.10)
(10) Tns:Past V.T:smell \ Tns:Past V.T:smell
(2125) recognizes nominal expression: 
(11) N.Cnt:cat Nmr:P1 \ Nom:(Head:(N.Cnt:cat Nmr:P1)) 
(12) N.Mass:lunch Nmr:Sg \ Nom:(Head:(N.Mass:lunch Nmr:Sg))

(2131) recognizes specific noun phrase with 'the': 
(13) Det:Def:the Nmr:P1 Num.Card:two Nmr:P1 Nom:(Head: 
(N.Cnt:cat Nmr:P1)) \ NP:(the two Head:(cat Nmr::P1))

(2137) recognizes specific noun phrase w/o determiner: 
(14) Nom:(Head:(N.Mass:lunch Nmr:Sg)) \ NP:(Head:(lunch Nmr:Sg))

(2173) recognizes passive: 
(15) V.Be:be en \ Pass

(2202) recognizes verbal with passive: 
(16) Pass V.T:prepare \ Vbl:(Pass V.T:prepare)

(2250) recognizes participial phrase: 
(17) ing Vbl:(Pass V.T:prepare) \ PartP:(ing Vbl:(Pass V.T:prepare))

(2500) recognizes participial phrase as noun attribute: 
(18) NP:(Head:(lunch Nmr:Sg)) PartP:(ing Vbl:(Pass V.T:prepare)) \ NP:(Head:(lunch Nmr:Sg) Nattr.PtP: 
(ing Vbl:(Pass V.T:prepare)))

(2201) recognizes verbal with transitive verb: 
(19) V.T:smell NP:(Head:(lunch Nmr:Sg) Nattr.PtP:(ing Vbl:(Pass V.T:prepare))) \ Vbl:(V.T:smell NP:(Head: 
(lunch Nmr:Sg) Nattr.PtP:(ing Vbl:(Pass V.T:prepare)))))

(2300) recognizes verb phrase: 
(20) Vbl:(V.T:smell NP:(Head:(lunch Nmr:Sg) Nattr.PtP:(ing Vbl:(Pass V.T:prepare)))) \ VP:(Vbl:(V.T:smell NP: 
(Head:(lunch Nmr:Sg) Nattr.PtP:(ing Vbl:(Pass V.T:prepare)))))

(2400) recognizes predicate: 
(21) Tns:Past VP:(Vbl:(V.T:smell NP:(Head:(lunch Nmr:Sg) 
Nattr.PtP:(ing Vbl:(Pass V.T:prepare))))) \ Pred: 
(Tns:Past VP:(Vbl:(V.T:smell NP:(Head:(lunch Nmr:Sg) 
Nattr.PtP:(ing Vbl:(Pass V.T:prepare)))))

The output from the Lexicon is:
NP:(the two Head:(cat Nmr::P1)) Pred:(Tns:Past VP:(Vbl: 
(V.T:smell NP:(Head:(lunch Nmr:Sg) Nattr.PtP:(ing Vbl: 
(Pass V.T:prepare))))))
ENTER SEMICON:

(3925) recognizes noun phrase as participant: (5.2)
(22) NP:(the two Head:(cat Nmr:P1)) \ P:(the two Head:(cat P1))
(23) NP:(Head:(lunch Nmr:Sg) Nattr.PtP:(ing Vbl:(Pass V.T: prepare))) \ P:(Head:(lunch Sg) Pattr.PtP:(ing Vbl: (Pass V.T:prepare)))

(3940) recognizes participial phrase as clause: (5.2)
(24) P:(Head:(lunch Sg) Pattr.PtP:(ing Vbl:(Pass V.T: prepare))) \ P:(Head:(lunch Sg) Pattr.PtP:(P:(Head: (lunch Sg)) Fred:(Time:Pres Vbl:(Pass V.T:prepare))))

(3950) recognizes passive clause: (5.6)
(25) P:(Head:(lunch Sg)) Pred:(Time:Pres Vbl:(Pass V.T: prepare)) \ Cl:(Time:Pres Rel:prepare(<animal><thing>) P1:(Ø $animal$ Ø) Focus:(P2:(Head:(lunch $thing$ Sg))))

(3992) recognizes declarative clause: (5.7)
(26) Cl:(Time:Pres Rel:prepare(<animal><thing>) P1:(Ø $animal$ Ø) Focus:(P2:(Head:(lunch $thing$ Sg)))) \ DeclCl:(Time:Pres Rel:prepare P1:(Ø $animal$ Ø) Focus:(P2:(Head:(lunch Sg))))

(3800) recognizes time: (5.3)
(27) Tns:Past VP:(V.T:smell P:(Head:(lunch Sg) Pattr.PtP:(DeclCl:(Time:Pres Rel:prepare P1:(Ø $animal$ Ø) Focus:(P2:(Head:(lunch Sg)))))) \ Time:Past Vbl:(V.T:smell P:(Head:(lunch Sg) Pattr.PtP:(DeclCl:(Time:Pres Rel:prepare P1:(Ø $animal$ Ø) Focus:(P2:(Head:(lunch Sg))))))

(3960) recognizes active clause with two participants: (5.6)
(28) P:(the two Head:(cat P1)) Pred:(Time:Past Vbl: (V.T:smell P:(Head:(lunch Sg) Pattr.PtP:(DeclCl: (Time:Pres Rel:prepare P1:(Ø $animal$ Ø) Focus:(P2:(Head:(lunch Sg)))))) \ Cl:(Time:Past Rel:smell(<animal><thing>) Focus:(P1:(the two Head: (cat $animal$ P1)) P2:(Head:(lunch $thing$ Sg) Pattr.PtP:(DeclCl:(Time:Pres Rel:prepare P1:(Ø $animal$ Ø) Focus:(P2:(Head:(lunch Sg)))))) \ DeclCl:(Time:Past Rel:smell Focus:(P1: (the two Head:(cat P1)) P2:(Head:(lunch Sg) Pattr.PtP:(DeclCl:(Time:Pres Rel:prepare P1:(Ø $animal$ Ø) Focus:(P2:(Head:(lunch Sg))))))

(3992) recognizes declarative clause: (5.7)
(29) Cl:(Time:Past Rel:smell(<animal><thing>) Focus:(P1: (the two Head:(cat $animal$ P1)) P2:(Head:(lunch $thing$ Sg) Pattr.PtP:(DeclCl:(Time:Pres Rel:prepare P1:(Ø $animal$ Ø) Focus:(P2:(Head:(lunch Sg)))))) \ DeclCl:(Time:Past Rel:smell Focus:(P1: (the two Head:(cat P1)) P2:(Head:(lunch Sg) Pattr.PtP:(DeclCl:(Time:Pres Rel:prepare P1:(Ø $animal$ Ø) Focus:(P2:(Head:(lunch Sg))))))


In line (25) of the trace, the passive clause does not have a P1 (cf. 5.6), so the variables for the head element and its number are instantiated to Ø, but the variable for the participant type, shown between dollar signs, is instantiated to that presupposed by the relation.

A diagram of the trace follows.

The clause is now built and the co-occurrence restrictions satisfied. As mentioned in 6.0, the task that remains is dependent on two things: first, the application in effect for the system at the time of processing, and second, the type of clause decoded.

In 2.5, two major applications for the present system are discussed. One use is as a translation device, the other as a knowledge base to be edited and queried. As indicated in that discussion, a decoding to the clause structure level is sufficient for rote translation purposes. Therefore, the clause structure just built (cf. (29) in the parse) would be passed to the transfer dictionary (cf. 2.5), and processing of the English would stop. If, however, the application were to edit and query a knowledge base, then
processing would have to continue. This is where the second matter mentioned above, that of clause type decoded, comes into consideration.

The clause type decoded in the parsing of 'The two cats smelled lunch being prepared' is declarative. 6.2 states that a declarative clause is simply an instruction to add the information to the existing network. As each particular sememe decoded is an instance of a more general class sememe (cf.1.2 or 6.2), it is assigned a number, generated by the system, to distinguish it from other instances of the same class sememe and to keep track of it for future reference. For example, the sememe 'cat' that is present in the Semicon is for the class 'cat'. The two cats mentioned in the sentence are specific instances of that more general concept, and so must be shown as such. The system generates and assigns numbers to each 'cat' such that there is 'cat-1' and 'cat-2' (cf.6.2). The event sememes 'smell' and 'prepare' in this clause are also particular instances of the class sememes 'smell' and 'prepare'. A number is also assigned to them by the system to differentiate these instances from other instances.

By binding these particular participants with these particular instances of events, the system tracks the fact that 'cat-1' and 'cat-2' 'smell(ed)-1' 'lunch-1', and that 'lunch-1' was 'prepare(d)-1' by some unspecified 'animal-1' (cf. parse line (25) above). This information is tested against currently existing information in the knowledge base
(cf. 6.2), and, if found compatible, stored for future reference.

When all of this has been done, we may say that the system has "understood" the sentence (cf. 6.2). We could test its understanding by asking it questions about the information it now has stored.

The parse of the next declarative sentence shows an indefinite determiner, a complex copula, a predicate nominative construction, and a prenominal adjective phrase.

DECODE: Some candy sounds like a pretty good idea.

ENTER MORPHICON:

(1) (1227) s o m e \ Det.Indef.Mass:some Nmr:Sg (3.2)
(2) (1082) c a n d y \ N.Mass:candy (3.1)
(3) (1982, 1446) s o u n d s \ C:3-Sg Tns:Pres V.C:sound (3.8)
(4) (1965) l i k e \ Prep:like (3.7)
(5) (1220) a \ Det.Indef.Cnt:a Nmr:Sg (3.2)
(6) (1943) p r e t t y \ Intns:pretty (3.6)
(7) (1705) g o o d \ Adj.Subj:good (3.5)
(8) (1036) i d e a \ N.Cnt:idea (3.1)

The output string from the Morphicon is:

ENTER LEXICON:

(2001) recognizes an complex copula: (4.20)
(9) V.C:sound Prep:like \ V.C.Cx:sound-like

(2100) assigns number to noun:
(10) N.Mass:candy \ N.Mass:candy Nmr:Sg
(11) N.Cnt:idea \ N.Cnt:idea Nmr:Sg

(2110) assigns tense to verb: (fn.1)
(12) Tns:Pres V.C:sound \ Tns:Pres V.C:sound

(2120) recognizes adjective phrase:
(13) Intns:pretty Adj.Subj:good \ AP:(Intns:pretty Adj.Subj:good)
(2125) recognizes nominal expression: (4.4)
\[ \text{N.Mass:candy Nmr:Sg} \setminus \text{Nom: (Head: (N.Mass:candy Nmr:Sg))} \]
\[ \text{AP: (Intns:pretty Adj.Subj:good) N.Cnt:idea Nmr:Sg} \setminus \]
\[ \text{Nom: (Modif: (Intns:pretty Adj.Subj:good) Head: (N.Cnt:idea Nmr:Sg))} \]

(2133) recognizes specific noun phrase: (4.5)
\[ \text{Det.Indef.Mass:some Nmr:Sg Nom: (Head: (N.Mass:candy Nmr:Sg)) \setminus NP: (some Head: (candy Nmr:Sg))} \]

(2132) recognizes specific noun phrase: (4.5)
\[ \text{Det.Indef.Cnt:a Nmr:Sg Nom: (Modif: (Intns:pretty Adj.Subj:good) Head: (N.Cnt:idea Nmr:Sg)) \setminus NP: (a Modif: (Intns: pretty Adj.Subj:good) Head: (idea Nmr:Sg))} \]

(2166) recognizes subject verb agreement: (4.13)
\[ \text{NP: (some Head: (candy Nmr:Sg)) C:3-Sg Tns:Pres \setminus NP: (some Head: (candy Nmr:Sg)) Tns:Pres} \]

(2204) recognizes verbal with complex copula and predicate nominative: (4.16)
\[ \text{V.Cx:sound-like NP: (a Modif: (Intns: pretty Adj.Subj:good) Head: (idea Nmr:Sg)) \setminus Vbl: (V.Cx:sound-like NP: (a Modif: (Intns: pretty Adj.Subj:good) Head: (idea Nmr:Sg)))} \]

(2502) recognizes compatibility of plurality of subject noun phrase and predicate nominative: (4.16)
\[ \text{NP: (some Head: (candy Nmr:Sg)) Tns:Pres Vbl: (V.Cx:sound-like NP: (a Modif: (Intns: pretty Adj.Subj:good) Head: (idea Nmr:Sg))) \setminus NP: (some Head: (candy Nmr:Sg)) Tns:Pres Vbl: (V.Cx:sound-like NP: (a Modif: (Intns: pretty Adj.Subj:good) Head: (idea Nmr:Sg)))} \]

(2300) recognizes verb phrase: (4.18)
\[ \text{Vbl: (V.Cx:sound-like NP: (a Modif: (Intns: pretty Adj.Subj:good) Head: (idea Nmr:Sg))) \setminus VP: (Vbl: (V.Cx:sound-like NP: (a Modif: (Intns: pretty Adj.Subj:good) Head: (idea Nmr:Sg))))} \]

(2400) recognizes predicate: (4.19)
\[ \text{Tns:Pres VP: (Vbl: (V.Cx:sound-like NP: (a Modif: (Intns: pretty Adj.Subj:good) Head: (idea Nmr:Sg)))) \setminus Pred: (Tns:Pres VP: (Vbl: (V.Cx:sound-like NP: (a Modif: (Intns: pretty Adj.Subj:good) Head: (idea Nmr:Sg))))}) \]

The output string from the Lexicon is:
\[ \text{NP: (some Head: (candy Nmr:Sg)) Pred: (Tns:Pres VP: (Vbl: (V.Cx:sound-like NP: (a Modif: (Intns: pretty Adj.Subj:good) Head: (idea Nmr:Sg)))))} \]
ENTER SEMICON:

(3800) recognizes time: (5.3)
(23) Tns:Pres VP:(Vbl:(V.C.Cx:sound-like NP:(a Modif:
        (Intns:pretty Adj.Subj:good) Head:(idea Nmr:Sg)))) \ 
        Time:Pres Vbl:(V.C.Cx:sound-like NP:(a Modif:
        (Intns:pretty Adj.Subj:good) Head:(idea Nmr:Sg)))

(3825) recognizes semantic interface of adjective: (5.2)
(24) Modif:(Intns:pretty Adj.Subj:good) \ PMODIF:
        (Intns:pretty good $subjective$)

(3830) semantic interface of adjective phrase: (5.2)
(25) PMODIF:(Intns:pretty good) \ PMODIF:(pretty $intensity$ $good$)

(3925) recognizes noun phrase as participant: (5.2)
(26) NP:(some Head:(candy Nmr:Sg)) \ P:(some Head:(candy
        Sg))
(27) NP:(a PMODIF:(pretty good) Head:(idea Nmr:Sg)) \ P:(a
        PMODIF:(pretty good) Head:(idea Sg))

(3970) recognizes active clause with complex copula and
        predicate nominative: (5.6)
(28) P:(some Head:(candy Sg)) Pred:(Time:Pres Vbl:
        (V.C.Cx:sound-like P:(a PMODIF:(pretty good) Head:
        (idea Sg)))) \ CL:(Time:Pres Rel:sound-like(<thing>
        <thing>) Focus:(P1:(some Head:(candy $thing$ Sg)))
        P2:(a PMODIF:(pretty good) Head:(idea $thing$ Sg)))

(3990) recognizes declarative clause: (5.7)
(29) CL:(Time:Pres Rel:sound-like(<thing><thing>) Focus:
        (P1:(some Head:(candy $thing$ Sg))) P2:(a PMODIF:
        (pretty good) Head:(idea $thing$ Sg))) \ Dec1CL:
        (Time:Pres Rel:sound-like Focus:(P1:(some Head:
        (candy Sg))) P2:(a PMODIF:(pretty good) Head:(idea
        Sg)))

The last example is a declarative clause which is
passive, is without participant one, has a locative expres-
sion, has modality realized as a probability expression, and
has perfective aspect.

DECODE: A pie had probably been eaten by the water.

ENTER MORPHICON:

(1) (1220) a \ Det.Indef.Cnt:a Nmr:Sg (3.2)
(2) (1027) p i e \ N.Cnt:pie (3.1)
(3) (1417) h a d \ Tns:Past Aux:have (3.8)
(4) (1881) probably \ Adv.Prob:probably  (3.6)  
(5) (1402) been \ en V.Be:be  (3.8)  
(6) (1462) eaten \ en V.T:eat  (3.8)  
(7) (1962) by \ Prep:by  (3.7)  
(8) (1200) the \ Det.Def:the Nmr:Sg  (3.2)  
(9) (1054) water \ N.Mass:water  (3.1)  

The output string from the Morphicon is:
Det.Indef.Cnt:a Nmr:Sg N.Cnt:pie Tns:Past Aux:have
Nmr:Sg N.Mass:water

ENTER LEXICON:

(2100) assigns number to noun:  (4.2)  
(10) N.Cnt:pie \ N.Cnt:pie Nmr:Sg  
(11) N.Mass:water \ N.Mass:water Nmr:Sg

(2110) assigns tense to verb: (fn.1)  (4.10)  
(12) Tns:Past Aux:have \ Tns:Past Aux:have

(2125) recognizes nominal expression:  (4.4)  
(13) N.Mass:water Nmr:Sg \ Nom:(Head:(N.Mass:water Nmr:Sg))  
(14) N.Cnt:pie Nmr:Sg \ Nom:(Head:(N.Cnt:pie Nmr:Sg))

(2131) recognizes specific noun phrase:  (4.5)  
(15) Det.Def:the Nmr:Sg Nom:(Head:(N.Mass:water Nmr:Sg)) \ 
        NP:(the Head:(water Nmr:Sg))

(2132) recognizes specific noun phrase:  (4.5)  
(16) Det.Indef.Cnt:a Nmr:Sg Nom:(N.Cnt:pie Nmr:Sg)) \ NP:(a 
        Head:(pie Nmr:Sg))

(2165) recognizes anataxis:  (4.12)  
(17) Tns:Past Aux:have Adv.Prob:probably en \ Tns:Past 
        Aux:have en Adv.Prob:probably

(2171) recognizes perfect:  (4.14)  
(18) Aux:have en \ Perf

(2173) recognizes passive:  (4.14)  
(19) V.Be:be en \ Pass

(2202) recognizes verbal with passive:  (4.16)  
(19) Pass V.T:eat\ Vbl:(Pass V.T:eat)

(2300) recognizes verb phrase:  (4.18)  
(20) Perf Adv.Prob:probably Vbl:(Pass V.T:eat) \ VP:(Perf 

(2180) recognizes a prepositional phrase:  (4.9)  
(21) Prep:by NP:(the Head:(water Nmr:Sg)) \ PP:(by NP:(the 
        Head:(water Nmr:Sg))
(2400) recognizes predicate: (4.19)
(22) Tns:Past VP: (Perf Adv.Prob:probably Vbl: (Pass V.T:eat) PP: (by NP: (the Head: (water Nmr:Sg)) \ Pred: (Tns:Past VP: (Perf Adv.Prob:probably Vbl: (Pass V.T:eat)) PP: (by NP: (the Head: (water Nmr:Sg)))))

The output string from the Lexicon is:
NP: (a Head: (pie Nmr:Sg)) Pred: (Tns:Past VP: (Perf Adv.Prob:probably Vbl: (Pass VT:eat)) PP: (by NP: (the Head: (water Nmr:Sg))))

ENTER SEMICON:

(3800) recognizes time: (5.3)

(3925) recognizes noun phrase as participant: (5.2)
(24) NP: (a Head: (pie Nmr:Sg)) \ P: (a Head: (pie Sg))
(25) NP: (the Head: (water Nmr:Sg)) \ P: (the Head: (water Sg))

(3876) recognizes semantic interface of locative expression: (5.4)
(26) PP: (by P: (the Head: (water Sg))) \ Loc: (by <thing> P: (the Head: (water $thing$ Sg)))

(3904) recognizes semantic interface of probability expression:
(27) Adv.Prob:probably \ Prob:probably $probability$

(3930) recognizes modality:
(28) Prob:probably \ Modality: (Prob:probably)

(3950) recognizes passive clause:
(28) P: (a Head: (pie Sg)) Pred: (Modality: (Prob:probably)
 Time: (Past Past) Vbl: (Pass V.T:eat) Loc: (by P: (the Head: (water Sg)))) \ Cl: (Modality: (Prob:probably)
 Time: (Past Past) Rel:eat (<animate> <concrete-object>)
 P1: (Ø $animate$ Ø) Focus: (P2: (a Head: (pie $concrete-object$ Sg))) Loc: (by P: (the Head: (water Sg))))

(3990) recognizes declarative clause:
(29) Cl: (Modality: (Prob:probably) Time: (Past Past) Rel:eat (<animate> <concrete-object>)
 P1: (Ø $animate$ Ø) Focus: (P2: (a Head: (pie $concrete-object$ Sg))) Loc: (by P: (the Head: (water Sg)))) \ DeclCl: (Modality: (Prob:probably) Time: (Past Past) Rel:eat P1: (Ø $animate$ Ø) Focus: (P2: (a Head: (pie Sg))) Loc: (by P: (the Head: (water Sg))))
7.2 Yes-no questions

In this section are declaratives with interrogative polarity. The first example to be parsed shows the so-called inversion of the subject noun phrase and the verb 'be', future tense expressed as 'be going to', and an objective case pronoun.

DECODE: Am I going to want them?

ENTER MORPHICON:

(1) (1404) a m \ C:1-Sg Tns:Pres V.Be:be  \\
(2) (1100) I \ Pro:1 Nmr:Sg  \\
(3) (1986,1550) go i n g \ ing V.I:go (fn.2)  \\
(4) (1906) t o \ Prep:to  \\
(5) (1466) w a n t \ V.T:want  \\
(6) (1113) t h e m \ Pro:3 Nmr:Pl m  \\

The output string from the Morphicon is:
C:1-Sg Tns:Pres V.Be:be Pro:1 Nmr:Sg ing V.I:go Prep:to
V.T:want Pro:3 Nmr:Pl m

ENTER LEXICON:

(2110) assigns tense to verb: (fn.1)  \\
(7) Tns:Pres V.Be:be \ Tns:Pres V.Be:be  \\

(2126) recognizes nominal expression:  \\
(8) Pro:1 Nmr:Sg \ Nom:(Head:(Pro:1 Nmr:Sg))  \\
(9) Pro:3 Nmr:Pl \ Nom:(Head:(Pro:3 Nmr:Pl))

(2134) recognizes specific noun phrase for pronoun:  \\
(10) Nom:(Head:(Pro:3 Nmr:Pl)) \ NP:(Head:(3 Nmr:Pl))  \\
(11) Nom:(Head:(Pro:1 Nmr:Sg)) \ NP:(Head:(1 Nmr:Sg))

(2163) recognizes ananaxis of subject and verb 'be': (4.11)  \\
(13) C:1-Sg Tns:Pres V.Be:be NP:(Head:(1 Nmr:Sg)) ing V.I:go
\ Q NP:(Head:(1 Nmr:Sg)) C:1-Sg Tns:Pres V.Be:be ing
V.I:go

(2168) recognizes subject verb agreement:  \\
(14) NP:(Head:(1 Nmr:Sg)) C:1-Sg Tns:Pres V.Be:be \ 
 NP:(Head:(1 Nmr:Sg)) Tns:Pres V.Be:be

(2172) recognizes progressive:  \\
(15) V.Be:be ing \ Prog

(2180) recognizes future tense:  \\
(16) Prog V.I:go Prep:to V.T:want \ Fut V.T:want
verb and locative adverb.

DECODE: Doesn't the man who ate bees live nearby?

ENTER MORPHICON:

(1) (1987, 1411) do es n't \ C:3-Sg Tns:Pres Aux:do not (fn.2) (3.8)
(2) (1200) the \ Det.Def:the Nmr:Sg (3.2)
(3) (1025) man \ N.Cnt:man (3.1)
(4) (1150) who \ Relv Pro:who (3.3)
(5) (1461) ate \ Tns:Past V.T:eat (3.8)
(6) (1980, 1021) bees \ N.Cnt:bee Nmr:Pl (fn.2) (3.1)
(7) (1553) live \ V.I:live (3.8)
(8) (1803) nearly \ Adv.Loc:nearby (3.6)

The output string from the Morphicon is:

ENTER LEXICON:

(2110) assigns tense to verb: (fn.1) (4.10)
(9) Tns:Pres Aux:do \ Tns:Pres Aux:do
(10) Tns:Past V.T:eat \ Tns:Past V.T:eat

(2100) assigns number to nouns:
(11) N.Cnt:man \ N.Cnt:man Nmr:Sg
(12) N.Cnt:bee Nmr:Pl \ N.Cnt:bee Nmr:Pl (fn.1)

(2126) recognizes nominal expression:
(13) N.Cnt:man Nmr:Sg \ Nom:(Head:(N.Cnt:man Nmr:Sg))
(14) N.Cnt:bee Nmr:Pl \ Nom:(Head:(N.Cnt:bee Nmr:Pl))

(2134) recognizes specific noun phrase with 'the':
(15) Det.Def:the Nmr:Sg Nom:(Head:(N.Cnt:man Nmr:Sg)) \ NP:(the Head:(man Nmr:Sg))

(2141) recognizes generic noun phrase for plural count nouns:
(16) Nom:(Head:(N.Cnt:bee Nmr:Pl)) \ NP:(Gen Head:(bee Nmr:Pl))

(2150) recognizes a relative pronoun as noun phrase and assigns it the number of its antecedent:
(17) NP:(the Head:(man Nmr:Sg)) Relv Pro:who \ NP:(the Head: (man Nmr:Sg)) Relv NP:(Head:(who Nmr:Sg))

(2200) recognizes a verbal with transitive verb:
(18) V.T:eat NP:(Gen Head:(bee Nmr:Pl)) \ Vbl:(V.T:eat NP: (Gen Head:(bee Nmr:Pl)))
(2300) recognizes a verb phrase:  
(19) Vbl:(V.T:eat NP:(Gen Head:(bee Nmr:P1))) \ VP:(Vbl: 
(V.T:eat NP:(Gen Head:(bee Nmr:P1))))

(2400) recognizes a predicate:  
(20) Tns:Past VP:(Vbl:(V.T:eat NP:(Gen Head:(bee Nmr:P1)))) 
\ Pred:(Tns:Past VP:(Vbl:(V.T:eat NP:(Gen Head:(bee 
Nmr:P1)))))

ENTER SEMICON:

(3800) recognizes time:  
(21) Tns:Past VP:(Vbl:(V.T:eat NP:(Gen Head:(bee Nmr:P1)))) 
\ Time:Past Vbl:(V.T:eat NP:(Gen Head:(bee Nmr:P1)))

(3815) recognizes semantic relations between relative 
pronoun and its antecedent:  
(22) NP:(the Head:(man Nmr:Sg)) Relv NP:(Head:(who Nmr:Sg)) 
\ NP:(the Head:(man $human$ Nmr:Sg)) Relv NP:(Head: 
(who $human$ Nmr:Sg))

(3925) recognizes noun phrase as participant:  
(23) NP:(Head:(who Nmr:Sg)) \ P:(Head:(who Sg)) 
(24) NP:(Gen Head:(bee Nmr:P1)) \ P:(Gen Head:(bee P1))

(3960) recognizes active clause with two participants:  
(25) P:(Head:(who Sg)) Pred:(Time:Past Vbl:(V.T:eat P:(Gen 
Head:(bee P1)))) \ Cl:(Time:Past Rel:eat(<animate> 
<concrete-obj>) Focus:(P1:(who $animate$ Sg)) P2:(Gen 
Head:(bee $concrete-object$ P1)))

(3991) recognizes relative clause:  
(26) Relv Cl:(Time:Past Rel:eat(<animate><concrete-obj>) 
Focus:(P1:(who $animate$ Sg)) P2:(Gen Head:(bee 
$concrete-object$ P1))) \ Relv Cl:(Time:Past Rel:eat 
Focus:(P1:(who Sg)) P2:(Gen Head:(bee P1)))

RE-ENTER LEXICON:

(2600) recognizes relative clause as a noun attribute:  
(27) NP:(the Head:(man Nmr:Sg)) Relv Cl:Time:Past Rel:eat 
Focus:(P1:(who Sg)) P2:(Gen Head:(bee P1))) \ NP:(the 
Head:(man Nmr:Sg) Nattr.RC:(Time:Past Rel:eat 
Focus:(P1:(who Sg)) P2:(Gen Head:(bee P1))))

(2161) recognizes anataxis of subject and aux 'do':  
(28) C:3-Sg Tns:Pres Aux:do not NP:(the Head:(man Nmr:Sg) 
Nattr.RC:(Time:Past Rel:eat Focus:(P1:(who Sg)) P2: 
(Gen Head:(bee P1)))) V.I:live \ Q NP:(the Head:(man 
Nmr:Sg) Nattr.RC:(Time:Past Rel:eat Focus:(P1:(who 
Sg)) P2:(Gen Head:(bee P1)))) not C:3-Sg Tns:Pres 
Aux:do V.I:live
(2168) recognizes subject verb agreement: (4.13)
(29) NP:(the Head:(man Nmr:Sg) Nattr.RC:(Time:Past Rel:eat Focus:(P1:(who Sg)) P2:(Gen Head:(bee P1)))) not C:3-Sg Tns:Pres \ NP:(the Head:(man Nmr:Sg) Nattr.RC:(Time:Past Rel:eat Focus:(P1:(who Sg)) P2:(Gen Head:(bee P1)))) not Tns:Pres

(2170) recognizes negative: (4.15)
(30) NP:(the Head:(man Nmr:Sg) Nattr.RC:(Time:Past Rel:eat Focus:(P1:(who Sg)) P2:(Gen Head:(bee P1)))) not Tns:Pres Aux:do \ NP:(the Head:(man Nmr:Sg) Nattr.RC:(Time:Past Rel:eat Focus:(P1:(who Sg)) P2:(Gen Head:(bee P1)))) Neg Tns:Pres

(2200) recognizes verbal with intransitive verb: (4.16)
(31) V.I:live \ Vbl:(V.I:live)

(2300) recognizes verb phrase: (4.18)
(32) Vbl:(V.I:live) \ VP:(Vbl:(V.I:live))

(2400) recognizes predicate: (4.19)
(33) Neg Tns:Pres VP:(Vbl:(V.I:live)) Adv.Loc:nearby \ Pred:

The output from the Lexicon is:
Q NP:(the Head:(man Nmr:Sg) Nattr.RC:(Time:Past Rel:eat Focus:(P1:(who Sg)) P2:(Gen Head:(bee P1)))) Pred:(Neg Tns:Pres VP:(Vbl:(V.I:live)) Adv.Loc:nearby)

RE-ENTER SEMICON:

(3925) recognizes noun phrase as participant: (5.2)
(34) NP:(the Head:(man Nmr:Sg) Nattr.RC:(Time:Past Rel:eat Focus:(P1:(who Sg)) P2:(Gen Head:(bee P1)))) \ P:(the Head:(man Sg) Partr.RC:(Time:Past Rel:eat Focus:(P1:(who Sg)) P2:(Gen Head:(bee P1))))

(3800) recognizes time: (5.3)
(35) Tns:Pres VP:(Vbl:(etc.)) \ Time:Pres Vbl:(etc.) (fn.3)

(3900) recognizes semantic interface of adverb: (5.4)
(36) Adv.Loc:nearby \ Loc:nearby $locative$

(3955) recognizes active clause with one participant: (5.6)
(37) P:(the Head:(man Sg) Partr.RC:(Time:Past Rel:eat Focus:(P1:(who Sg)) P2:(Gen Head:(bee P1)))) Pred:
    (Neg Time:Pres Vbl:(V.I:live) Loc:nearby) \ Neg C1:(Time:Pres Rel:live$<$animal$> Focus:(P1:(the Head:
        (man $animal$ Sg) Partr.RC:(Time:Past Rel:eat Focus:(P1:(who Sg)) P2:(Gen Head:(bee P1))))
    Loc:nearby)
(3992) recognizes declarative clause with interrogative and negative polarity:

(38) Q Neg Cl: (Time:Pres Rel:live<animal> Focus:(P1:(the Head:(man $animal$ Sg) Pattr.RC:(Time:Past Rel:eat Focus:(P1:(who Sg)) P2:(Gen Head:(bee P1))))) Z 1 1
Loc:nearby) \ Polarity:(Q Neg) DeclCl: (Time:Pres Rel:live Focus:(P1:(the Head:(man Sg) Pattr.RC:(Time:Past Rel:eat Focus:(P1:(who Sg)) P2:(Gen Head:(bee P1))))) Loc:nearby)

7.3 Wh-questions

The first example parse shows an idiomatic verb, an interrogative pronoun, a proper noun, and negative polarity.

DECODE: Whom didn't Rudolph look in on?

ENTER MORPHICON:

(1) (1151) w h o m .. \ Intr..Pro:who m
(2) (1967,1412) d i d n ' t \ Tns:Past Aux:do not(fn.2)
(3) (1001) R u d o l p h \ N.Prop:Rudolph
(4) (1448) l o o k \ V.I:look
(5) (1961) i n \ Prep:in
(6) (1964 o n \ Prep:on

The output string from the Morphicon is:


ENTER LEXICON:

(2110) assigns tense to verb: (fn.1)
(7) Tns:Past Aux:do \ Tns:Past Aux:do

(2100) assigns number to nouns: (fn.1)
(8) N.Prop:Rudolph \ N.Prop:Rudolph Nmr:Sg

(2003) recognizes idiom:
(9) V.I:look Prep:in Prep:on \ V.T:look-in-on

(2125) recognizes nominal expression:
(10) N.Prop:Rudolph Nmr:Sg \ Nom:(Head:(N.Prop:Rudolph Nmr:Sg))

(2130) recognizes specific noun phrase with proper noun:
(11) Nom:(Head:(N.Prop:Rudolph Nmr:Sg)) \ NP:(Head:(Rudolph Nmr:Sg))
(2155) recognizes a interrogative pronoun as noun phrase and assigns it singular number:  
   (4.7) 
   (12) Intr..Pro:who \ Intr..NP:(Head:(who Nmr:Sg)) 

(2161) recognizes anataxis of subject and aux 'do':  
   (4.11) 
   (13) Tns:Past Aux:do not NP:(Head:(Rudolph Nmr:Sg))  
       V.T:look-in-on \ Q NP:(Head:(Rudolph Nmr:Sg)) not  
       Tns:Past Aux:do V.T:look-in-on 

(2170) recognizes negative:  
   (4.15) 
   (14) NP:(Head:(Rudolph Nmr:Sg)) not Tns:Past Aux:do \  
       NP:(Head:(Rudolph Nmr:Sg)) Neg Tns:Past 

(2201) recognizes verbal with transitive verb:  
   (4.16) 
   (15) V.T:look-in-on NP:(Head:(who Nmr:Sg)) m \ Vbl:  
       (V.T:look-in-on NP:(Head:(who Nmr:Sg)))) 

(2300) recognizes verb phrase:  
   (4.18) 
   (16) Vbl:(V.T:look-in-on NP:(Head:(who Nmr:Sg)))) \ VP:(Vbl:  
       (V.T:look-in-on NP:(Head:(who Nmr:Sg)))) 

(2400) recognizes predicate:  
   (4.19) 
   (17) Neg Tns:Past VP:(Vbl:(V.T:look-in-on NP:(Head:(who  
         Nmr:Sg)))) \ Pred:(Neg Tns:Past VP:(Vbl:  
         (V.T:look-in-on NP:(Head:(who Nmr:Sg)))))) 

The output from the Lexicon is:  
   Intr.Q NP:(Head:(Rudolph Nmr:Sg)) Pred:(Neg Tns:Past VP:(Vbl:  
   (V.T:look-in-on NP:(Head:(who Nmr:Sg)))) 

(The two dots in the output from the Morphicon are shown to  
be instantiated to "Q NP Neg Tns V.T" in the above output  
from the Lexicon.)

ENTER SEMICON: 

(3925) recognizes noun phrase as participant:  
   (5.2) 
   (18) NP:(Head:(Rudolph Nmr:Sg)) \ P:(Head:(Rudolph Sg))  
   (19) NP:(Head:(who Nmr:Sg)) \ P:(Head:(who Sg)) 

(3800) recognizes time:  
   (5.3) 
   (20) Tns:Past VP:(Vbl:(etc.)) \ Time:Past Vbl:(etc.) (fn.3) 

(3960) recognizes active clause with two participants:  
   (5.6) 
   (21) P:(Head:(Rudolph Sg)) Pred:(Neg Time:Past  
       Vbl:(V.T:look-in-on P:(Head:(who Sg)))) \ Neg  
       Cl:(Time:Past Rel:look-in-on<animal><animal>) Focus:  
       (P1:(Head:(Rudolph $animal$ Sg))) P2:(Head:(who  
       $animal$ Sg)))
(3990) recognizes interrogative clause with negative polarity:

(22) Intr Q Neg Cl:(Time:Past Rel:look-in-on(<animal> <animal>) Focus:(P1:(Head:(Rudolph $animal$ Sg)))
P2:(Head:(who $animal$ Sg))) Polarity:Neg IntrCl:
(Time:Past Rel:look-in-on Focus:(P1:(Head:(Rudolph Sg))) P2:(Head:(who Sg)))

The second example trace shows an interrogative adverb, perfect and progressive aspect, and the spelling rule for the suffix -ing.

DECODE: Where has he been taking naps?

ENTER MORPHICON:

(1) (1921) where \ Intr..Adv.Loc:where
(2) (1416) have \ C:3-Sg Tns:Pres Aux:have
(3) (1103) he \ Pro:3-male Nmr:Sg
(4) (1402) been \ N. Be:be
(5) (1986,1463) taken \ V.T:take (fn.2)
(6) (1980,1035) naps \ N.Cnt:nap Nmr:P1

The output string from the Morphicon is:
Intr..Adv.Loc:where C:3-Sg Tns:Pres Aux:have Pro:3-male Nmr:Sg en V.Be:be ing V.T:take N.Cnt:nap Nmr:P1

ENTER LEXICON:

(2110) assigns tense to verb: (fn.1)
(7) Tns:Pres Aux:have \ Tns:Pres Aux:have

(2100) assigns number to nouns: (fn.1)
(8) N.Cnt:nap Nmr:P1 \ N.Cnt:nap Nmr:P1

(2125) recognizes nominal expression:
(9) N.Cnt:nap Nmr:P1 \ Nom:(Head:(N.Cnt:nap Nmr:P1))

(2126) recognizes nominal expression:
(10) Pro:3-male Nmr:Sg \ Nom:(Head:(Pro:3-Sg Nmr:Sg))

(2136) recognizes specific noun phrase with plural count noun:
(11) Nom:(Head:(N.Cnt:nap Nmr:P1)) \ NP:(Head:(nap Nmr:P1))

(2134) recognizes specific noun phrase with pronoun:
(12) Nom:(Head:(Pro:3-male Nmr:Sg)) \ NP:(Head:(3-male Sg))
(2162) recognizes anataxis of subject and aux 'have': (4.11)
(13)  C:3-Sg Tns:Pres Aux:have NP:(Head:(3-male Nmr:Sg)) en
V.Be:be \ Q NP:(Head:(3-male Nmr:Sg)) C:3-Sg Tns:Pres
Aux:have en V.Be:be

(2166) recognizes subject verb agreement: (4.13)
(14)  NP:(Head:(3-male Nmr:Sg)) C:3-Sg Tns:Pres Aux:have \ 
NP:(Head:(3-male Nmr:Sg)) Tns:Pres

(2201) recognizes verbal with transitive verb: (4.16)
(15)  V.T:take NP:(Head:(nap Nmr:Pl)) \ Vbl:(V.T:take 
NP:(Head:(nap Nmr:Pl)))

(2171) recognizes perfect: (4.14)
(16)  Aux:have en \ Perf

(2172) recognizes progressive: (4.14)
(17)  V.Be:be ing \ Prog

(2300) recognizes verb phrase: (4.18)
(18)  Perf Prog Vbl:(V.T:take NP:(Head:(nap Nmr:Pl)))) \ VP:
(Perf Prog Vbl:(V.T:take NP:(Head:(nap Nmr:Pl))))

(2400) recognizes predicate: (4.19)
(19)  Tns:Pres VP:(Perf Prog Vbl:(V.T:take NP:(Head:(nap 
Nmr:Pl)))) \ Adv.Loc:where \ Pred:(Tns:Pres VP:(Perf 
Prog Vbl:(V.T:take NP:(Head:(nap Nmr:Pl))))
Adv.Loc:where)

The output from the Lexicon is:
Intr Q NP:(Head:(3-male Nmr:Sg)) Pred:(Tns:Pres VP:(Perf 
Prog Vbl:(V.T:take NP:(Head:(nap Nmr:Pl)))) Adv.Loc:where)
(The two dots in the output from the Morphicon are shown to
be instantiated to "Q NP Tns Perf Prog V.T NP" in the above
output from the Lexicon.)

ENTER SEMICON:

(3925) recognizes noun phrase as participant: (5.2)
(20) NP:(Head:(3-male Nmr:Sg)) \ P:(Head:(3-male Sg))
(21) NP:(Head:(nap Nmr:Pl)) \ P:(Head:(nap Pl))

(3800) recognizes time: (5.3)
(22) Tns:Pres VP:(Perf Prog Vbl:(etc.))3 \ Time:(Pres 
Past Pres) Vbl:(etc.)

(3811) recognizes pronoun as pronominal form: (fn.4) (5.2)
(23) Head:(3-male Sg) \ Proform:(Head:(Max $male$ Sg))

(3900) recognizes semantic interface of adverb: (5.4)
(3960) recognizes active clause with two participants: (5.6) 
(25) P: (Head: (Max Sg)) Pred: (Time: (Pres Past Pres)) Vbl: 
(V.T: take P: (Head: (nap P1))) \ C1: (Time: (Pres Past 
Pres)) Rel: take(<thing><thing>) Focus: (P1: (Head: (Max 
$thing$ Sg))) P2: (Head: (nap $thing$ P1))

(3990) recognizes interrogative clause: (5.7) 
(22) Intr Q C1: (Time: (Pres Past Pres)) Rel: take(<thing> 
<thing>) Focus: (P1: (Head: (Max $thing$ Sg))) P2: (Head: 
(nap $thing$ P1)) \ IntrCl: (Time: (Pres Past Pres) 
Rel: take Focus: (P1: (Head: (Max Sg))) P2: (Head: (nap 
P1)))
NOTES

Chapter 7

1. This general rule is applied because there is a match. In this particular case, however, the (local) output is the same as the (local) input. See the actual (not instantiated) rules and their description in Chapter 3 - 5 for further explanation.

2. These rules have the same number in the trace because, before the procedure gives the output from the first rule, it performs the other(s). These other rules verify whether the co-occurrence restrictions stated in the first rule are met.

3. 'Etc.' is used here for the sake of brevity.

4. We supply an antecedent here since there is no context which does.
Chapter 8

CONCLUSION

8.0 Conclusion

The objective of this dissertation, to show that the linguistic information that a typical speaker possesses can be uniformly represented as a stratified system of signs, has been met. In Chapters 3, 4, and 5, it is shown how information can be encapsulated in a semiotic format with four strata of linguistic structure. Furthermore, Chapters 6 and 7 discuss how the linguistic information is used to produce and understand texts, how the system as a whole is integrated.

Although the body of this dissertation addresses basic linguistic information, I have confidence that more complex information can be handled by the same technique as that elaborated here; that is, by representing it as a sign. Linguistic knowledge is the linking of some content and its expression, and that relationship can be stated for all information, no matter what its complexity.

In this treatment, I tried to adhere to a dictionary system which was strictly stratified, as first proposed in Outline of Stratificational Grammar. In the process of stating the linguistic information, however, I found a strictly stratified accounting to be cumbersome and uneconomical in some ways. One way in which strict stratification is inefficient is in the separation of morphotactic and
lexotactic information. By definition, only graphemic and
morphemic information belongs in the Morphicon, but, as
explained in 2.1 and 3.0, I chose to partially coalesce
lexemic with morphemic information in the Morphicon and
state lexotactic sub-categories of morphologically defined
natural classes. Another uneconomical burden of strict
stratification, which I nonetheless adhered to, is the
separate statements of morphological, lexicological, and
semological information for "regular" forms. In effect,
separate dictionary entries in the Morphicon and Semicon had
to be made for the sake of stratification when a single
entry, with all of the information, might as well have
sufficed. Given these findings, we might do well in the
future to question the advisability of pursuing a strictly
stratified model.
APPENDIX A
MORPHICON

(1001) Rudolph \ N.Prop:Rudolph
(1002) Houston \ N.Prop:Houston
(1003) February \ N.Prop:February
(1004) Monday \ N.Prop:Monday
(1005) Christmas \ N.Prop:Christmas
(1020) cat \ N.Cnt:cat
(1021) bee \ N.Cnt:bee
(1022) story \ N.Cnt:story
(1023) woman \ N.Cnt:woman
(1024) women \ N.Cnt:woman Nmr:Pl
(1025) man \ N.Cnt:man
(1026) men \ N.Cnt:man Nmr:Pl
(1027) pie \ N.Cnt:pie
(1028) car \ N.Cnt:car
(1029) evening \ N.Cnt:evening
(1030) middle \ N.Cnt:middle
(1031) mitten \ N.Cnt:mitten
(1032) box \ N.Cnt:box
(1033) wife \ N.Cnt:wife
(1034) wives \ N.Cnt:wife Nmr:Pl
(1035) nap \ N.Cnt:nap
(1036) idea \ N.Cnt:idea

(1050) honey \ N.Mass:honey
(1051) work \ (N.Mass, V.T, V.I):work
(1052) power \ N.Mass:power
(1053) money \ N.Mass:money
(1054) water \ N.Mass:water

(1080) lunch \ (N.Cnt, N.Mass):lunch
(1081) time \ (N.Cnt, N.Mass):time
(1082) candy \ (N.Cnt, N.Mass):candy

(1100) I \ Pro:1 Nmr:Sg
(1101) we \ Pro:1 Nmr:Pl
(1102) you \ Pro:2 Nmr:(Sg,Pl) (Ø,m)
(1103) he \ Pro:3-male Nmr:Sg
(1104) she \ Pro:3-female Nmr:Sg
(1105) it \ Pro:3-neuter Nmr:Sg (Ø,m)
(1106) they \ Pro:3 Nmr:Pl
(1107) me \ Pro:1 Nmr:Sg m
(1108) us \ Pro:1 Nmr:Pl m
(1110) him \ Pro:3-male Nmr:Sg m
(1111) her \ Pro:3-female Nmr:Sg m
(1113) them \ Pro:3 Nmr:Pl m

(1150) who .. \ (Relv, Intr) .. Pro:who
(1151) whom .. \ (Relv, Intr) .. Pro:who m
(1152) which .. \ (Relv, Intr) .. Pro:which
(1153) that .. \ (Relv .. Pro:that, Det.Def:that
Nmr:Sg ..)
what \\

the \ (Det.Def, Det.Gen):the Nmr:(Sg,Pl)
(1201) this \ Det.Def: this Nmr:Sg
(1202) these \ Det.Def: this Nmr:Pl
(1203) those \ Det.Def: that Nmr:Pl

(1220) a \ (Det.Indef.Cnt, Det.Gen):a Nmr:Sg
(1221) every \ Det.Indef.Cnt: every Nmr:Sg
(1222) each \ Det.Indef.Cnt: each Nmr:Sg
(1223) either \ Det.Indef.Cnt: either Nmr:Sg
(1224) neither \ Det.Indef.Cnt: neither Nmr:Sg
(1225) much \ Det.Indef.Mass: much Nmr:Sg
(1226) many \ Det.Indef.Cnt: many Nmr:Pl
(1227) some \ (Det.Indef.Mass: some Nmr:Sg,
    Det.Indef.Cnt: some Nmr:Pl)
(1228) any \ (Det.Indef.Mass: any Nmr:Sg,

(1400) be \ V.Be:be
(1401) being \ (ing V.Be:be, N.Cnt:being)
(1402) been \ en V.Be:be
(1403) is \ C:3-Sg Tns:Pres V.Be:be
(1404) am \ C:1-Sg Tns:Pres V.Be:be
(1405) are \ Tns:Pres V.Be:be
(1406) was \ C:(3-Sg,1-Sg) Tns:Past V.Be:be
(1407) were \ Tns:Past V.Be:be

(1410) do \ (Aux, V.T, V.I):do
(1411) does \ C:3-Sg Tns:Pres (Aux, V.T, V.I):do
(1412) did \ Tns:Past (Aux, V.T, V.I):do
(1413) done \ en (V.T, V.I):do

(1415) have \ (Aux, V.T):have
(1416) has \ C:3-Sg Tns:Pres (Aux, V.T):have
(1417) had \ (Tns:Past (Aux, V.T):have, en V.T:have)

(1420) will \ Mod:will (Ø, Tns:Fut)
(1421) won't \ Mod:will (Ø, Tns:Fut) not
(1422) can \ (N.Cnt:can, V.T:can, Mod:can)
(1423) can't \ Mod:can not
(1424) could \ (Mod:can Tns:Past, Mod:could)
(1425) should \ Mod:should

(1440) become \ (Ø, en) V.C:become
(1441) became \ Tns:Past V.C:become
(1442) seem \ V.C:seem
(1443) feel \ (V.C, V.T):feel
(1444) felt \ (Tns:Past, en) (V.C, V.T):feel
(1445) act \ (V.C, V.T):act
(1446) sound \ (V.C, V.T):sound
(1447) taste \ (V.C, V.T):taste
(1448) look \ (V.C, V.T):look
(1449) smell \ (V.C, V.T, V.I):smell
(1460) eat \ T.V:eat
(1461) ate \ Tns:Past V.T:eat
(1462) eaten \ en V.T:eat
(1463) take \ T.V:take
(1464) took \ Tns:Past V.T:take
(1465) taken \ en V.T:take
(1466) want \ V.T:want
(1467) put \ (Ø, Tns:Past, en) V.T:put
(1468) run \ (Ø, en) (V.T, V.I):run
(1469) ran \ Tns:Past (V.T, V.I):run
(1470) prepare \ V.T:prepare
(1471) wear \ V.T:wear
(1472) wore \ Tns:Past V.T:wear
(1473) worn \ en V.T:wear

(1550) go \ V.I:go
(1551) gone \ en V.I:go
(1552) went \ Tns:Past V.I:go
(1553) live \ V.I:live
(1554) sit \ V.I:sit
(1555) sat \ Tns:Past V.I:sit
(1556) sleep \ (V.I, N.Cnt):sleep
(1557) slept \ (Tns:Past, en) V.I:sleep

(1600) one \ Num.Card:one Nmr:Sg
(1601) 1 \ Num.Card:one Nmr:Sg
(1602) two \ Num.Card:two Nmr:Pl
(1603) 2 \ Num.Card:two Nmr:Pl
(1604) three \ Num.Card:three Nmr:Pl

(1650) first \ Num.Ord:first
(1651) second \ Num.Ord:second
(1652) third \ Num.Ord:third

(1700) sick \ Adj.Subj:sick
(1701) long \ Adj.(Subj, Obj):long
(1702) happy \ Adj.Subj:happy
(1703) cold \ (Adj.(Subj, Obj), N.(Cnt,Mass)):cold
(1704) funny \ Adj.Subj:funny
(1705) good \ Adj.Subj:good
(1706) superior \ Adj.Subj:superior
(1707) peculiar \ Adj.Subj:peculiar
(1720) round \ Adj.Obj:round
(1721) little \ Adj.(Subj, Obj):little
(1722) large \ Adj.(Subj, Obj):large
(1723) sad \ Adj.Subj:sad

(1740) old \ Adj.Age:old
(1741) young \ Adj.Age:young
(1742) new \ Adj.Age:new

(1760) red \ Adj.Col:red
(1761) black \ Adj.Col:black
(1762) orange \ (Adj.Col, N.Cnt):orange

(1780) metallic \ Adj.Clas:metallic
(1781) woolen \ Adj.Clas:woolen
(1782) domestic \ Adj.Clas:domestic
(1783) electric \ Adj.Clas:electric

(1800) here \ Adv.Loc:here
(1801) there \ Adv.Loc:there
(1802) outside \ Adv.Loc:outside
(1803) nearby \ Adv.Loc:nearby

(1820) nearly \ Adv.Degr:nearly
(1821) almost \ Adv.Degr:almost
(1822) practically \ Adv.Degr:practically

(1840) now \ Adv.Temp:now
(1841) then \ Adv.Temp:then
(1842) today \ Adv.Temp:today
(1843) formerly \ Adv.Temp:formerly

(1844) o'clock \ Adv.Temp.Hr:o'clock

(1860) always \ Adv.Freq:always
(1861) sometimes \ Adv.Freq:sometimes
(1862) never \ Adv.Freq:never
(1863) often \ Adv.Freq:often
(1864) rarely \ Adv.Freq:rarely

(1880) certainly \ Adv.(Prob, Intns):certainly
(1881) probably \ Adv.Prob:probably
(1882) possibly \ Adv.Prob:possibly

(1900) rapidly \ Adv.Manr:rapidly
(1901) cautiously \ Adv.Manr:cautiously
(1902) respectfully \ Adv.Manr:respectfully
(1903) easily \ Adv.Manr:easily

(1920) how .. \ Intr .. Adv.Manr:how
(1921) where .. \ Intr .. Adv.Loc:where
(1922) when .. \ Intr .. Adv.Temp:when

(1940) very \ Intns:very
(1941) really \ (Intns, Adv.Prob):really
(1942) quite \ Intns:quite
(1943) pretty \ (Adj.Subj, Intns):pretty

(1950) not \ not

(1960) to \ Prep:to
(1961) in \ Prep:in
(1962) by \ Prep:by
(1963) at \ Prep:at
(1964) on \ Prep:on
(1965)  like \ (Prep:like, N.Cnt:like, V.T:like)


(1981)  N.X:Y<-s, -z, -x, -ch, -sh, -y/i>  -es \  
        N.X:Y  Nmr:Pl

(1982)  V.X:Y  -s \ C:3-Sg  Tns:Pres  V.X:Y

(1983)  V.X:Y<-s, -z, -x, -ch, -sh, -y/i>  -es \  
        C:3-Sg  Tns:Pres  V.X:Y

(1984)  V.X:Y  -d \  (Tns:Past, en)  V.X:Y

(1985)  V.X: (Y, Y<-y/i>, Y<-X1/X1X1>)  -ed \  
       (Tns:Past, en)  V.X:Y
       IF X1 = (b,d,g,p,t,m,n,l,r)

(1986)  V.X: (Y, Y<-e/Ø>, Y<-X1/X1X1>)  -ing \  
       ing V.X:Y  IF X1 = (b,d,g,p,t,m,n,l,r)

(1987)  (Aux:X, Mod:Y [Tns:X1], V.Be:be)  -n't \  
       (Aux:X, Mod:Y [Tns:X1], V.Be:be)  not
LEXICON

(2001) V.C:X Prep:like \ V.C.Cx:X-like
(2002) V.T:take .. Prep:on \ V.T:take-on ..
(2030) N.Cnt:cat N.Cnt:nap \ N.Cnt:cat-nap
(2031) Adj.Age:old N.Mass:money \ N.Mass:old-money
(2032) V.I:sit Prep:in \ N.Cnt:sit-in
(2033) Num.Card:X Nmr:Y Adv.Temp.Hr:o'clock \ N.Hour:X-o'clock IF X = (one,two,three,four,
five,six,seven,eight,nine,ten,eleven, twelve)
(2100) N.X:Y Nmr:(Pl,Ø) \ N.X:Y Nmr:(Pl,Sg)
(2111) Mod:X Tns:(Y,Ø) [not] \ Mod:X [not] Tns:(Y, Pres)

(2126) Pro:X Nmr:Y \ Nom:(Head:(Pro:X Nmr:Y))
(2130) [Num.Card:X Nmr:Y] Nom:(..Head:(N.Prop:X1 Nmr:Y) \ NP:([X]..Head:(X1 Nmr:Y))
    Nom:(..Head:(N.X2:Y2 Nmr:Y)) \ 
    NP:(X "([X1] [Y1])..Head:(Y2 Nmr:Y))

    Nom:(..Head:(N.Cnt:Y1 Nmr:Y)) \ 
    NP:(X [X1]..Head:(Y1 Nmr:Y))

(2133) Det.Indef.Mass:X Nmr:Sg Nom:(..Head:(N.Mass:Y Nmr:Sg)) \ NP:(X..Head:(Y Nmr:Sg))

(2134) Nom:(Head:(Pro:X Nmr:Y)) \ NP:(Head:(X Nmr:Y))

(2136) [Num.Card:X Nmr:P1] Nom:(..Head:(N.Cnt:Y Nmr:P1)) \ 
    NP:([X]..Head:(Y Nmr:P1))

(2137) Nom:(..Head:(N.(Mass,Hour):X Nmr:Sg)) \ 
    NP:(..Head:(X Nmr:Sg))

(2140) Det.Gen:X Nmr:Sg Nom:(..Head:(N.Cnt:Y Nmr:Sg)) \ 
    NP:(Gen..Head:(Y Nmr:Sg))

(2141) Nom:(..Head:(N.Cnt:X Nmr:P1)) \ 
    NP:(Gen..Head:(X Nmr:P1))

(2142) Nom:(..Head:(N.Mass:X Nmr:Sg)) \ 
    NP:(Gen..Head:(X Nmr:Sg))

(2150) NP:(..Head:(X Nmr:Y)) [m] Relv .. Pro:X1 \ 
    NP:(..Head:(X Nmr:Y)) [m]
    Relv .. NP:(Head:(X1 Nmr:Y)) [m]

(2155) Intr .. Pro:X \ Intr .. NP:(Head:(X Nmr:Sg))

      V.Y2:X3

      V.Y2:X3 \ Q NP:X1 [not] [C:X] Tns:Y Aux:have

      V.Y2:X3, AP:Y3, NP:X4)  \ Q NP:X1 [not] [C:X]
      Tns:Y V.Be:be ((en,ing) [Adv.Y1:X2] V.Y2:X3,

      [Adv.Y1:X2] (V.Y2:X3, Aux:have)

(2165) [C:X] Tns:Y (Aux:(do,have), V.Be:be) ^{([not]}
      [Adv.X1:Y1)*) [ing,en]  \ [not] [C:X] Tns:Y
      (Aux:(do,have), V.Be:be) [ing,en] [Adv.X1:Y1]

(2166) NP:(...)Head:(X Nmr:Sg)...) [not] [Adv.Y:X1] C:3-Sg
      Tns:Pres \ NP:(...)Head:(X Nmr:Sg)...) [not]
      Tns:Pres [Adv.Y:X1] IF X NOT = (1, 2)

(2167) Q NP:(Head:(1 Nmr:Sg)...) not [C:1-Sg] Tns:Pres
      V.Be:be \ Q NP:(Head:(1 Nmr:Sg)...) not
      Tns:Pres V.Be:be

(2168) NP:(Head:(1 Nmr:Sg)...) [not] C:1-Sg Tns:X V.Be:be \ 
      NP:(Head:(1 Nmr:Sg)...) [not] Tns:X V.Be:be

(2169) NP:(...)Head:(X Nmr:Sg)...) [not] C:3-Sg Tns:Past
      V.Be:be \ NP:(...)Head:(X Nmr:Sg)...) [not]
      Tns:Past V.Be:be IF X NOT = (1,2)
(2170) NP:X (Mod:Y not Tns:X1, not Tns:X1 (V.Be:be, 
          Aux:(have,do))) \ NP:X (Neg Mod:Y Tns:X1, 
          Neg Tns:X1 (V.Be:be, Aux:(have,∅)))

(2171) Aux:have en \ Perf
(2172) V.Be:be ing \ Prog
(2173) V.Be:be en \ Pass

          V.X1:Y1 \ Fut [Perf] [Prog] [Pass] [Adv.X:Y]
          V.X1:Y1

(2180) Prep:X NP: (.Head: (Y Nmr:X1) ..)
         (∅,m IF Y = (1,2,3,3-Y1,who)) \ 
         PP: (X NP: (.Head: (Y Nmr:X1) ..))

(2200) V.I:X \ Vbl: (V.I:X)

(2201) V.T:X NP: (.Head: (Y Nmr:X1) ..)
         (∅,m IF Y = (1,2,3,3-Y2,who)) \ 
         Vbl: (V.T:X NP: (.Head: (Y Nmr:X1) ..))

         IF X NOT = Loc

(2203) (V.Be:be, V.C:X) [Adv.Y:X1] AP:Y1 \ 

(2204) (V.Be:be, V.C:become, V.C.Cx:X) [Adv.Y:X1] NP:Y1 \ 
         NP:Y1)

(2250) ing [Perf] Vbl:X \ PartP: (ing [Perf] Vbl:X)

(2300) [[Perf] [Fut [Perf]]] [Prog] [Adv.X:Y] Vbl:X1 \ 
       VP: ([[Perf] [Fut [Perf]]] [Prog] [Adv.X:Y]
       Vbl:X1) IF X NOT = Loc
\[(2400) \ [\text{Neg}] \ [\text{Mod}:X] \ Tns:Y \ VP:X1 \sim ([\text{Adv}.Y1:X2] [\text{PP}:Y2])^* \ \backslash \ \\
\quad \text{Pred}:([[\text{Neg}] \ [\text{Mod}:X] \ Tns:Y \ VP:X1 \sim ([\text{Adv}.Y1:X2] [\text{PP}:Y2])^*)\]
\[(2500) \ NP:X \ [m] \ \text{PartP}:Y \ \backslash \ NP:(X \ \text{Nattr.} \ \text{PtP}:Y) \ [m] \]
\[(2501) \ NP:X \ [m] \ \text{PP}:Y \ \backslash \ (NP:(X \ \text{Nattr.} \ \text{PP}:Y) \ [m], \ NP:X \ [m] \ \text{PP}:Y)\]
\[(2502) \ NP:(..\text{Head}:(X \ \text{Nmr}:Y)\ldots) \ \ldots \ \text{Vbl}:(..(\text{V.Be}:be, \ \\
\quad \text{V.C}:\text{become, V.C}:Cx:X1) \ NP:(..\text{Head}:(Y1 \ \text{Nmr}:Y, \ \\
\quad \quad (\text{who, which, what}) \ \text{Nmr}:Sg))) \ \backslash \ \\
\quad NP:(..\text{Head}:(X \ \text{Nmr}:Y)\ldots) \ \ldots \ \text{Vbl}:(..(\text{V.Be}:be, \ \\
\quad \text{V.C}:\text{become, V.C}:Cx:X1) \ \\
\quad \quad NP:(..\text{Head}:(Y1,\text{who, which, what}) \ \text{Nmr}:Y)))\]
\[(2600) \ NP:X \ [m] \ \text{RelvCl}:Y \ \backslash \ NP:(X \ \text{Nattr.} \ \text{RC}:Y) \ [m]\]
SEMICON

(3001) eat \ ingest + etc. (animate <concrete-object>)
(3002) put \ do + etc. (animate <thing> <locative>)
(3003) walk \ go + (step + at the same time as) <animal>
(3004) be \ exist (<thing> <(thing, quality, locative)>
(3005) seem \ etc. (<thing> <(thing, quality)>

(3500) man \ human + male + adult + etc.
(3501) Houston \ place + city + in Texas + etc.
(3502) Monday \ day + etc.
(3503) X-o'clock \ temporal-point + etc.
(3504) story \ abstract-object + etc.
(3551) Christmas \ day + etc.
(3552) February \ time-period + etc.
(3553) car \ inanimate + etc.
(3554) water \ inanimate + etc.
(3555) here \ locative + etc.
(3556) always \ frequency + etc.
(3557) almost \ degree + etc.
(3558) now \ temporal + etc.

(3600) concrete-object \ thing + (potentially) tangible
(3601) abstract-object \ thing + intangible
(3602) place \ inanimate + locative
(3603) animate \ concrete-object + living
(3604) inanimate \ concrete-object
(3605) animal \ animate + etc.
(3606) plant \ animate
(3607) human \ animal + etc.
(3608) neuter \ -male + -female
(3609) emotion \ subjective
(3610) humor \ subjective
(3611) temporal \ abstract-object

(3650) old \ quality + age + etc. \ ant:(new,young)
(3651) black \ quality + color + etc.
(3652) sad \ quality + emotion + etc. \ ant:(happy,glad)
(3653) funny \ quality + humor + etc.

(3675) the \ definite
(3676) that \ definite + distal
(3677) this \ definite + proximal
(3680) a \ indefinite
(3685) who \ human
(3686) which \ -human
(3687) that \ thing

(3700) (Pl, plural) \ more than one
(3701) (Sg, singular) \ one

(3703) (Gen, generic) \ class + etc.
(3704) Proform \ definite + etc.
(3705) Speaker \ interlocutor + etc.
(3705) Hearer \ interlocutor + etc.

(3750) at \ locative <(temporal-point, place)>
(3751) on \ locative + surface <(concrete-object, day)>
(3752) in \ locative + interior <(concrete-object, abstract-object IF NOT (day Sg) AND temporal-point)>

(3753) by \ locative + proximity <concrete-object>

(3755) by \ end point in time <temporal>

(3754) to \ goal <(concrete-object, temporal)>

(3780) temporal-point \ temporal

(3781) time-period \ temporal

(3782) day \ temporal

(3790) very \ intensity + etc.

(3800) Tns:X VP: ([[Perf] [Fut [Perf]]] [Prog] ..) \ Time: (X [[Past] [Fut [Past]]] [Pres]) ..

(3810) Head:((1,2) X) \ Proform:(Head:((Speaker,Hearer) X))

(3811) Head:(3-X Sg) \ Proform:(Head:(Y $X$ Sg))

IF Y NOT = (Speaker, Hearer)

(3812) Head:(3 Pl) \ Proform:(Head:(X Pl))

IF X NOT = (Speaker,Hearer)

(3815) NP:(..Head:(X Nmr:Y)) [m] Relv ..

NP:(Head:(X1 Nmr:Y)) \ NP:(..Head:(X $Y1$ Nmr:Y)) [m] Relv ..

NP:(Head:(X1 $Y1$ Nmr:Y))


IF Y1 NOT = X1
(3825) (AP, Modif): (. . Adj. Subj: X) / 
    (PrAdj, PModif): (. . X $subjective$)
    (PrAdj, PModif): (. . X $(size, shape)$)
(3827) (AP, Modif): (. . Adj. Age: X) / 
    (PrAdj, PModif): (. . X $age$)
    (PrAdj, PModif): (. . X $color$)
(3829) (AP, Modif): (. . Adj. Clas: X) / 
    (PrAdj, PModif): (. . X $classifier$)
(3830) (PrAdj, PModif): (((X: Y, Intns: X1) Y1)*) / 
    (PrAdj, PModif): (((X, Y1 $intensity$) Y1)*)
    IF X NOT = Adv. X2

(3875) Patr.PP: (X P: (. . Head: (Y X1) . .)) / 
    Patr.PP: (Loc: (X<thing> P: (. . Head: (Y $thing$ X1) . .))
(3876) PP: (X P: (. . Head: (Y X1) . .)) / 
    Loc: (X<thing> P: (. . Head: (Y $thing$ X1) . .))
(3877) Patr.PP: (X P: (. . Head: (Y X1) . .)) / 
    Patr.PP: (Temp: (X<temporal> 
    P: (. . Head: (Y $temporal$ X1) . .))
(3878) PP: (X P: (. . Head: (Y X1) . .)) / 
    Temp: (X<temporal> P: (. . Head: (Y $temporal$ X1) . .))

(3900) Adv. Loc: X \ Loc: X $locative$
(3901) Adv. Temp: X \ Temp: X $temporal$
(3902) Adv. Degr: X \ Degr: X $degree$
Adv.Freq:X \ Freq:X $frequency$
(3904) Adv.Prob:X \ Prob:X $probability$
(3905) Adv.Manr:X \ Manr:X $manner$
(3925) NP:(..Head:(X Nmr:Y) [Nattr.X1:Y1]*) [X2] \n   P:(..Head:(X Y) [Pattr.X1:Y1]*) [X2]
   IF X2 NOT = Relv
(3930) [Mod:X]..[Prob:Y] \ Modality:([X] [Prob:Y]) ..
(3940) P:(..Head:(X Y)..Pattr.PtP:(ing [Perf] Vbl:X1)) \n   P:(..Head:(X Y)..Pattr.PtP:(P:(Head:(X Y))
   Pred:(Time:(Pres [Past]) Vbl:X1)))
(3950) P:(..Head:(X Y)...) Pred:([Neg] [Modality:X1] Time:X2
   [Y2:X3]* Vbl:(Pass V.T:Y3) ~([X4:(by P:(..Head:
   (Y4 X5)...) [Y5:X6]*)]) \ [Neg] C1:([Modality:X1]
   Time:X2 Rel:Y3(<Y6><X7>) P1:(..Head:(Y4 $Y6$
   X5)...) Focus:(P2:(..Head:(X $X7$ Y)...) [Y2:X3]* [Y5:X6]* [X4:(by P:(..Head:(Y4 X5)...) IF P1:(Y4 = ∅)])
(3955) P:(..Head:(X Y)...) Pred:([Neg] [Modality:X1]
   Time:X2 [Y2:X3]* Vbl:(V.I:Y3) [X4:Y4]*) \ [Neg]
   C1:([Modality:X1] Time:X2 Rel:Y3<X5> Focus:(P1:
   (..Head:(X $X5$ Y)...) [Y2:X3]* [X4:Y4]*)
(3960) P: (.Head: (X Y) ..) Pred: ([Neg] [Modality: X1]
  Time: X2 [Y2: X3]* Vbl: ((V.T: Y3 P: (.Head: (X4 Y4) ..)) [X5: Y5]*) \ [Neg] C1: ([Modality: X1]
  Time: X2 Rel: Y3 (<X6> <Y6>) Focus: (P1: (.Head: (X $X6$ Y) ..)) P2: (.Head: (X4 $Y6$ Y4) ..)
  [Y2: X3]* [X5: Y5]*)

(3965) P: (.Head: (X Y) ..) Pred: ([Neg] [Modality: X1]
  Time: X2 [Y2: X3]* Vbl: ((V.Be: be, V.C: Y3)
  PrAdj: (.X4)) [Y4: X5]*) \ [Neg] C1: ([Modality: X1]
  Time: X2 Rel: (be, Y3) (<Y5> <quality>)
  Focus: (P1: (.Head: (X $Y5$ Y) ..))
  P2: (.X4 $quality$) [Y2: X3]* [Y4: X5]*)

(3970) P: (.Head: (X Y) ..) Pred: ([Neg] [Modality: X1]
  Time: X2 [Y2: X3]* Vbl: ((V.Be: be, V.C: become,
  V.Cx: Y3) P: (.Head: (X4 Y4) ..)) [X5: Y5]*) \ [Neg] C1: ([Modality: X1] Time: X2
  Rel: (be, become, Y3) (<X6> <thing>)
  Focus: (P1: (.Head: (X $X6$ Y) ..))
  P2: (.Head: (X4 $thing$ Y4) ..) [Y2: X3]* [X5: Y5]*)

(3990) Intr [Q] [Neg] C1: X \ [Polarity: Neg] IntrCl: X


(3992) Ø [Q] [Neg] C1: X \ [Polarity: ([Q] [Neg])] DeclCl: X
APPENDIX B
ABBREVIATIONS

Symbols which have suffixes (i.e., words or abbreviations separated by a period) should be broken down to be looked up in this appendix. For example, for the tactic code V.C.Cx, the reader should look up V, then C, then Cx to determine that this code means "verb copula complex", or a complex copula.

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>zero, nil</td>
</tr>
<tr>
<td>1</td>
<td>first person</td>
</tr>
<tr>
<td>2</td>
<td>second person</td>
</tr>
<tr>
<td>3</td>
<td>third person</td>
</tr>
<tr>
<td>Adj</td>
<td>adjective</td>
</tr>
<tr>
<td>Adv</td>
<td>adverb</td>
</tr>
<tr>
<td>ant</td>
<td>antonym</td>
</tr>
<tr>
<td>AP</td>
<td>adjective phrase</td>
</tr>
<tr>
<td>Aux</td>
<td>auxiliary</td>
</tr>
<tr>
<td>C</td>
<td>copula if a suffix (V.C, V.C.Cx)</td>
</tr>
<tr>
<td>Cx</td>
<td>concord</td>
</tr>
<tr>
<td>Card</td>
<td>cardinal</td>
</tr>
<tr>
<td>Cl</td>
<td>clause</td>
</tr>
<tr>
<td>Clas</td>
<td>classifier</td>
</tr>
<tr>
<td>Col</td>
<td>color</td>
</tr>
<tr>
<td>Cnt</td>
<td>count</td>
</tr>
<tr>
<td>Cx</td>
<td>complex</td>
</tr>
<tr>
<td>DeclC</td>
<td>declarative clause</td>
</tr>
<tr>
<td>Def</td>
<td>definite</td>
</tr>
<tr>
<td>Det</td>
<td>determiner</td>
</tr>
<tr>
<td>en</td>
<td>past participle</td>
</tr>
<tr>
<td>Freq</td>
<td>frequency</td>
</tr>
<tr>
<td>Fut</td>
<td>future</td>
</tr>
<tr>
<td>Gen</td>
<td>generic</td>
</tr>
<tr>
<td>Hr</td>
<td>hour</td>
</tr>
<tr>
<td>I</td>
<td>intransitive</td>
</tr>
<tr>
<td>Indef</td>
<td>indefinite</td>
</tr>
<tr>
<td>ing</td>
<td>present participle</td>
</tr>
<tr>
<td>Intns</td>
<td>intensifier</td>
</tr>
<tr>
<td>Intr</td>
<td>interrogative</td>
</tr>
<tr>
<td>IntrC</td>
<td>interrogative clause</td>
</tr>
<tr>
<td>Loc</td>
<td>locative</td>
</tr>
<tr>
<td>m</td>
<td>objective case</td>
</tr>
<tr>
<td>Manr</td>
<td>manner</td>
</tr>
<tr>
<td>Mod</td>
<td>modal</td>
</tr>
<tr>
<td>Modif</td>
<td>modifier</td>
</tr>
<tr>
<td>N</td>
<td>noun</td>
</tr>
<tr>
<td>Nattr</td>
<td>noun attribute</td>
</tr>
<tr>
<td>Neg</td>
<td>negative</td>
</tr>
<tr>
<td>Nmr</td>
<td>number</td>
</tr>
<tr>
<td>Nom</td>
<td>nominal</td>
</tr>
<tr>
<td>NP</td>
<td>noun phrase</td>
</tr>
<tr>
<td>Num</td>
<td>numeral</td>
</tr>
</tbody>
</table>
Obj ............. objective
Ord ............. ordinal
P ................. participant
P1 ................. participant one
P2 ................. participant two
PartP ............. participial phrase
Pass .............. passive
Patr .............. participant attribute
Perf ............. perfect
Pl ................. plural
PModif ............ participant modifier
PP ................. prepositional phrase
PrAdj ............. predicate adjective
Pred ............. predicate
Prep ............. preposition
Pres ............. present
Pro .............. pronoun
Prob ............. probability
Proform ........... pronominal form
Prog ............. progressive
Prop ............. proper
PtP ............... participial phrase
Q ................. interrogative (question)
Rel ................ relation (event)
RC ................. relative clause
Relv ............. relative
RelvCl ............ relative clause
Sg ................ singular
Subj ................ subjective
T .................. transitive
Temp ............. temporal
Tns ................ tense
V .................. verb
Vbl ................. verbal
VP ................. verb phrase
X + number ....... variable
Y + number ....... variable
REFERENCES


Amsterdam: Bejamins Publishing Company.

Ikegami, Yoshihiko. 1969. The semological structure of the 
English verbs of motion. New Haven: Linguistic Automa-
tion Project, Yale University.

Johannesson, Nils-Lennart. 1976. English modal auxiliaries: 
a stratificational account. Stockholm: Almqquist and 
Wiksell International.

Kress, G.R., ed. 1976. Halliday: system and function in 

Lamb, Sydney M. 1966. Outline of stratificational grammar 
(revised). Washington, D.C.: Georgetown University 
Press.

Lamb, Sydney M. 1971. The crooked path of progress in cogni-
tive linguistics. In Georgetown University Round Table 
on Language and Linguistics (GURT), 1971, ed. by 
town University Press. Reprinted in Makkai and Lock-
wood 1973:12-33.

Lamb, Sydney M. 1984. Semiotics of language and culture: a 
relational approach. In The Semiotics of Culture and 
Language, ed. by Robin F.P. Fawcett, M.A.K. Halliday, 
Sydney M. Lamb, and Adam Makkai, 2.71-100. London and 
Dover, N.H.: Frances Pinter.

Lamb, Sydney M. 1985. Information and its representation in 
English texts. In Computing Power and Legal Reasoning, 
lishing Co.

information system. In Georgetown University Round 
Table on Languages and Linguistics, 1986, ed. by Simon 
University Press.

Lamb, Sydney M. unpublished ms. On defining the language. 
Read at the 1986 meeting of LACUS.

Lamb, Sydney M. forthcoming. On constructing the content of 
a text. To appear in Discourse Semantics and Text 
Semantics: Proceedings of the 1984 Rice Symposium on 
Linguistics and Semiotics, ed. by James E. Copeland.

ms.


