SOME HEURISTICS FOR AUTOMATIC DETECTION AND RESOLUTION OF ANAPHORA IN DISCOURSE

by

Sharon Baranofsky

NL-15 April 1973

Natural Language Research for CAI
Supported by
The National Science Foundation
Grant GJ 509X

The Department of Computer Sciences
and CAI Laboratory
The University of Texas
Austin, Texas 78712

This paper was submitted to the Graduate School in satisfaction of requirements for the Master of Arts degree.

ABSTRACT

Computer processing of a natural language discourse involves translating the discourse into internal information structures which contain relations between elements of the discourse. These internal information structures represent a meaning interpretation for the discourse. The structures must contain relations that form connections between elements of the discourse, such as the relation between an anapheric expression and its antecedent, if accuracy in automatic question answering and paraphrasing is desired. This paper presents the design of an anaphora resolution subunit of a natural language processor that includes heuristics for detecting the following types of anaphora in a discourse and associating them with their antecedents: third person personal pronouns, anaphoric noun phrases introduced by demonstrative adjectives, relative pronouns, and anaphoric noun phrases marked by the definite determiner 'the.'

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I: INTRODUCTION

Computer processing of a discourse written in a natural language involves automatically translating the strings of natural language symbols in the discourse into information structures which make explicit certain relations that exist between the symbols of the discourse. A set of relations existing between the symbols of a discourse may be considered a meaning interpretation for the discourse. The task for which a discourse is being processed indicates the set of relations (or the meaning interpretation) which must be considered in deriving the transformed information structures.

For example, if the task, or application, of a language processor is question answering, then the translated structures must contain certain relations which form inter-sentential connections. Otherwise, the application area of the language processor would be limited to those questions for which the answers can be derived from a single analyzed, or processed, sentence. Each sentence of a connected discourse is dependent on the conceptual framework previously developed in the discourse. The relationships existing between the symbols of a sentence and symbols of the preceding sentences of the disourse must be contained in the transformed information structures in order to establish this dependency. One inter-sentential relation

which, if made explicit in the transformed information structures, would extend the question answering capabilities of a language processor beyond the intra-sentence range is the relation between an anaphoric expression and its antecedent.

According to John Olney (1967):

> Essentially, an expression is anaphoric if: 1) it is understood in context in a sense more specific than can be obtained as a compositional function of its lexical sense(s) and those of words directly related to it syntactically (surface structure); and 2) its specialized sense can be obtained at least in part as a compositional function of the senses just referred to and the sense of some preceding expression (its correct antecedent).

Olney (1967) emphasizes the importance of associating anaphoric expressions in a connected discourse with their antecedents:

> Until a capability is developed for associating anaph ric expressions with the expressions they replace, computer processing of English text will largely be confined within sentence boundaries. That is, computer programs will be unable to exploit syntactic and other structural features for appropriately relating the information expressed in one sentence of a text with that expressed by other sentences of the same text. This confinement within sentence boundaries places a rather low upper bound on the accuracy of computer processing of natural language text for applications in which analysis of the specific information content of the text is required, as for example 'fact' retrieval or auto-abstracting.

This paper presents the design of an anaphora resolution subunit of a natural language processor. This design embodies some heuristic procedures based on an analysis of some patterns of anaphora which occur frequently in English discourse and a study of English pronominalization and discourse referencing. The reader may acquire a sufficient linguistic background

to understand the heuristics presented by referring to the work of Karttunen, Lakoff, and Ross.

Some heuristics are presented for automatically detecting the following types of anaphoric expression in a discourse and associating them with their antecedents:

i) third person personal pronouns

ii) noun phrases modified by a demonstrative adjective

iii) relative pronouns

iv) noun phrases marked with the definite determiner 'the'

The heuristic procedures are stated precisely, in chapters 4 through 7, in the form of LISP functions, which are dependent on the set of primitives described in the Appendix. Although the LISP functions are described in detail in chapters 4 through 7, it is assumed that the reader has some knowledge of LISP 1.5, as given in McCarthy (1968). The reader should refer to the Appendix for a description of any primitive called in the LISP functions whose operation is not clear.

CHAPTER II: SOME COMPONENTS OF A LANGUAGE PROCESSOR

The heuristics for anaphora resolution are dependent on some components of a natural language processor. These components, which are described below, are a lexicon, a semantic structure into which input sentences are translated, a list of concept pairs, and a list of main topics. The lexicon to be described is based on the current lexical structure for Simmons' experimental language processor. The translated semantic structure, which includes Fillmore's case structure notion, is the structure into which input sentences are transformed by Simmons' processor.

IIa. The Lexicon

The lexicon to be described contains syntactic and definitional information for each word in an English discourse to be processed. This information is used by the recognizer and transformational component of the processor to determine if a string of English words input into the system is an acceptable string according to the specifications of the grammar, and to translate an acceptable input string into internal information structures.

Each entry, or word, in the lexicon has associated with it all of its word senses. For example, an entry for the word

'palm' might have three word senses associated with it. One sense of 'palm' would refer to the tropical tree, the second would refer to the inner surface of the hand, and the third sense of 'palm' would refer to the act of hiding something in the hand.

Each of the word senses of a lexical entry has certain syntactic and definitional properties associated with it. These properties are stored in lists of attribute-value pairs. The attribute word class is contained on the syntactic property list associated with each word sense in the lexicon. The syntactic property list associated with a noun also includes the attribute's number, gender, and semantic marker. The syntactic property list associated with a pronoun includes the attribute's number, gender, and person. The syntactic properties associated with verbs, adjectives, adverbs, and prepositions are not given here since this information is not relevant to resolution of the anaphoric constructions considered here.

Some of the definitional attributes of a word sense, placed on a definitional property list associated with the word sense, are:

SUP, which indicates a superset or inclusion relation with another lexical entry word sense.

SUB, which indicates a subset relation with another entry word sense.

TYPE, which specifies any peculiarities which distinguish the

entry word sense from other entry word senses having the same SUP relation (i.e. from other members of the same concept class).

EQUIV, which indicates an equivalence with another entry word sense in the lexicon.

HAS-PART, which specifies a list of one or more entry word senses which are part(s) of the word sense in question.

PART-OF, which specifies an entry word sense of which the entry word sense in question is part.

Some or all of the above definitional attributes may be associated with a particular entry word sense. Figure 1 shows the content of the lexical entry for the word 'palm.'

Word Senses	Syntactic and Definitional Properties
palm1	SYNT (WC noun) (NBR sing) (GEND neuter) (PERSON 3) (MKR tree1)
	DEF (SUP tree1) (TYPE tropical2) (HAS-PART fronds1)
palm2	SYNT (WC noun) (NBR sing) (GEND neuter) (PERSON 3)
	DEF (SUP surface2) (TYPE skin1) (PART-OF hand2)
palm3	SYNT (WC verb)
	DEF (SUP hide1) (TYPE in hand2)

Figure 1. Sample Lexical Entry

IIb. The Translated Information Strustures

The process by which an English input string is recognized and translated into an internal information structure need

not be described here. However, the form of the translated structures will be described. The form of the translated structures parallels that of the lexical entries. Each constituent formed, according to the transformations specified in the grammar processor, from words of an input sentence, has associated with it lists of syntactic and definitional properties which have been derived from the lexical entries for the words that make up the constituent and from the natural language context of the constituent. Although the constituents formed in processing a sentence have syntactic property lists associated with them, the syntactic information is mainly used in the syntactic analysis of processing the sentence, and these lists are of little importance after the sentence has been processed. It is the definitional property lists of the constituents formed for a sentence which constitute the internal information structure into which the input sentence is translated.

The translated structures for a sentence input to the system are centered around the main verb(s) of the sentence. A verbal constituent is a primary translated structure for the sentence, with pointers to substructures which are in certain relationships with the verb. A verbal constituent contains such definitional information as:

i) A TOK relation with the lexical entry word sense of the verb (since the verb's usage in the sentence, restricted by its context, is a specific instance of the more general concept of the verb which is the lexical entry).

ii) The relations SUBJ and OBJ with noun phrase constituents formed for the sentence, which indicate the surface structure subject and object of the verb.

iii) Adverbial relations of MANNER, TIME, and LOC, that are specified in the input sentence by prepositional phrases and adverbs. MANNER, TIME, and LOC specify a manner of action, a time, and a location associated with the verb of the verbal constituent.

iv) Relations specified by the preposition of a prepositionsl phrase.

A noun phrase constituent may have the following relations on its definitional list:

TOK, which specifies that a token of the lexical entry word sense of the noun is used in forming the constituent.

ASSOC, which indicates an association between two nouns which are adjacent in the surface structure of an input sentence.

POSS, which indicates a possessive modifier of the noun phrase.

MOD, which indicates a modifier of the noun phrase.

S, which indicates a sentence modifier of the noun phrase.

Conjoined forms are handled by forming a constituent which has a TOK relation with the conjunction in the lexicon, and the relations EL1 and EL2 pointing to the elements of the conjoined form.

Each time a relation with a constituent or lexical entry is put on the definitional list of a constituent, the inverse relation is put on the definitional list of the constituent

or lexical entry. For example, if the relation (TOK protect2) is put on the definitional list of the constituent G8 (i.e. G8 = (TOK protect2)...), then the definitional list of the lexical entry word sense 'protect2' become (protect2 = (TOK-1 G8)...).

During processing of a discourse, a list is kept of the main verbal constituents formed in the order of their formation. This list consists of verbal constituents formed for simple sentences, verbal constituents formed for main clauses joined by co-ordinate conjunctions, and verbal constituents formed for adverbial clauses which are subordinate to a main verbal constituent. The verbal constituents formed for relative clauses are not included in this list of main verbal constituents. In Figure 2, G9 is a verbal constituent which would be added to the list of main verbal constituents.

Consider the translated structures of the input sentence given in figure 2. The syntactic properties of constituents are not given, only the definitional properties, which constitute the translated information structures.

INPUT SENTENCE: The seed, which is the coconut of commerce, is securely protected from the action of sea water by its thick fibrous husk and hard shell.

TRANSLATED STRUCTURE:

G1 = (TOK seed1)(DET the1)(EL1-1 G2)(SUBJ-1 G9)

G2 = (TOK is2)(EL1 which1 G1)(EL2 G3)

G3 = (TOK coconut1)(DET the1)(EL2 G2)(OF commerce1)

G4 = (TOK action1)(DET the1)(OF G5)(FROM-1 G9)

G5 = (TOK water1)(OF-1 G4)(ASSOC Sea2)

G6 = (TOK husk1)(EL1-1 G8)(MOD thick2 fibrous1)(POSS its1)

G7 = (TOK shell1)(EL2-1 G8)(MOD hard2)(POSS its1)

G8 = (TOK and1)(EL1 G6)(EL2 G7)(BY-1 G9)

G9 = (TOK protect2)(FROM G4)(BY G8)(MANNER securely1)(SUBJ G1)

Figure 2. Sample Translated Information Structures

IIc. The Set of Acceptable Concept Combinations

A list of acceptable concept combinations specifies the ways in which certain concepts may be or are often combined. It contains such members as (EAT FOOD), (ANIMAL SLEEP), (LIVING-BEING ACT), etc. Often, if the heuristics for deriving the antecedent of a particular anaphoric expression are applied, more than one antecedent candidate is found. If all candidates meet the syntactic specifications of agreement (in gender and number) with the anaphoric expression, then a semantic analysis of the antecedent candidates must be performed to decide which is the correct antecedent.

Semantic analysis of an antecedent candidate involves examining concepts which are in a SUP relation with the antecedent candidate and concepts which are in a SUP relation with the context of the anaphoric expression to see if a concept SUP to the antecedent candidate and a concept SUP to the context of the anaphoric expression occur on the list of acceptable concept

combinations. (The function CONCEPTCHK, described in the Appendix, performs such an examination.) The antecedent candidate which conveys a concept which can be acceptably combined with the conceptual context of the anaphoric expression is selected as the correct antecedent.

IId. The Ordered List of Main Topics

An ordered list of main topics is formed by putting each noun or noun phrase referenced in a discourse on a list as it is processed. A numerical weight is associated with each item on the list, and the list is ordered according to decreasing values of the weights. When an item is first put on the list, it is given an initial weight of one. Each time a noun or noun phrase is referenced, either directly or anaphorically, the list is searched to see if the item is already on the list. If it is, its weight is increased by one and the list is re-ordered. If not, it is put on the list and a weight of one is associated with it. Each time a sentence is processed, those items on the list which were not referenced in the sentence just processed have their weights decreased by 1/2. Items are removed from the list if their weights do not meet a prescribed minimum value after a number of sentences has been processed. In this way, the main topicsof a discourse, which are referenced throughout the discourse, remain at the front of the ordered list. Also, those topics which were most recently talked about in the discourse will be at the front of the list.

CHAPTER III: DETECTION OF ANAPHORA IN A DISCOURSE

Before the heuristics for deriving the antecedent of an anaphoric expression can be applied, the problem of detecting anaphoric expressions in an input sentence must be solved. The third person pronoun 'it,' the demonstrative adjectives, and the relative pronouns have more than one word sense associated with the entry for each in the lexicon. One of the word senses associated with each of these entries is the anaphoric sense. The recognition unit, or the grammar processor, of the language processor must be equipped with the appropriate contextual information to determine when an occurrence of one of these words in an input sentence corresponds to the anaphoric sense of the word. When the anaphoric sense of one of these words occurring in an input sentence is selected, according to the specifications in the grammar processor, that anaphoric expression is put on a list of anaphoric expressions to be resolved when processing of the sentence is completed.

The contextual information, which the grammar processor must contain in order to detect anaphoric senses of words of each of the types of anaphora being dealt with in this study, is discussed in the following sections.

IIIa. Detection of Third Person Pronouns

The third person personal pronouns are 'he,' 'him,'

'his,' 'she,' 'her,' 'hers,' 'it,' 'its,' 'they,' 'them,' 'their,' and 'theirs.' The pronoun 'it' has two word senses associated with it in the lexicon. One sense is the abstract sense, and the other the anaphoric sense. All of the other third person pronouns have only one word sense, the anaphoric sense. The word class associated with each anaphoric word sense of a third person pronoun in the lexicon is ANAPRO.

Assume that the recognizer of the language processor can detect cleft sentences, such as (1a), and transform them to the internal structures which correspond to (1b). Also assume that idiomatic expressions which involve the abstract pronoun 'it,' such as (2), can be detected by the recognizer and translated to meaningful internal structures. For syntactic and semantic patterns which constitute cleft sentences or idiomatic usages, the grammar specifies that the recognizer select the non-anaphoric abstract sense of 'it.' For an occurrence of 'it' in all other syntactic and semantic environments, the grammar specifies that the recognizer select the anaphoric sense.

1a) It worried John that he was sick.

1b) That he was sick worried John.

2) It rains.

When the word class ANAPRO occurs on the left-hand side of a rule in the grammar processor, the semantics associated with that rule specify that a copy of the particular instance

of anaphora which matched the rule and the constituent in which the anaphoric pronoun is placed be put on a list of anaphoric expressions to be resolved when processing of the sentence is completed. After the sentence has been translated to the internal information structure, each anaphoric expression which occurred in that sentence will be resolved by techniques (specified in chapters 5 - 8) which correspond to the type of anaphora.

IIIb. Detection of Demonstrative Adjectives

Since the demonstrative adjectives 'that' and 'those' are most often used deictically, rather than anaphorically, they will not be dealt with in this study. The demonstrative adjectives 'this' and 'these' followed by a noun or noun phrase in a discourse will be regarded as anaphoric. In those cases in which 'this' or 'these' followed by a noun or noun phrase is deictic, no antecedent should be found in the discourse by following the procedures for deriving the antecedent of an anaphoric noun phrase introduced by a demonstrative adjective.

The lexical entries for 'this' and 'these' have particular word senses with the word class DEMADJ on their syntactic property lists. 'This' and 'these' may also be used as demonstrative pronouns, but as demonstrative pronouns they are not followed by a noun or noun phrase. The grammar specifies that, for 'this' or 'these' followed by a noun or noun phrase, the word sense be selected whose word class is DEMADJ. The semantics associated with the rules in the grammar whose left-hand

sides are (DEMADJ + N) or (DEMADJ + NP) specify that a copy of the constituent formed for the noun phrase be placed on the list of anaphora to be resolved when processing of the sentence is completed.

IIIc. Detection of Relative Pronouns

The relative pronouns are 'who,' 'whom,' 'whose,' 'which,' and 'that.' The lexical entries for 'who,' 'whom,' 'whose,' and 'which' have two word senses associated with each of them. One sense is the interrogative adjective sense, and the other the relative pronoun sense. The entry for 'that' has four senses associated with it. These senses correspond to the usage of 'that' as a demonstrative pronoun, a demonstrative adjective, a conjunction introducing a noun clause, and a relative pronoun.

The rules in the grammar processor specify that the relative pronoun sense be selected for an occurrence of one of these symbols in an input sentence in the following syntactic environment:

i) preceded by a noun phrase or a clause, except in the case of 'which' and 'whom,' which may be preceded by a preposition and then preceded by a noun phrase,

ii) followed by a verb phrase or a clause introduced by a definite noun phrase (a name, a third person pronoun, a noun phrase with a determiner or a demonstrative adjective modifying it, or a plural noun phrase), except in the case

of 'whose,' which is followed by the noun or noun phrase it modifies and is then followed by either a verb phrase or a clause introduced by a definite noun phrase.

The specification that a relative pronoun may be preceded by a clause is made in the grammar in order to handle extraposed relative clauses. The procedures for deriving the antecedent of a relative pronoun apply for those sentences with extraposed relative clauses which are syntactically unambiguous. In (3) the antecedent of 'who' is syntactically (and semantically) ambiguous. (4) and (5) are not ambiguous.

3) A woman appeared before the judge who wanted to kill him.

4) A man went to the fair who lost his mind.

5) A man died who had been sick for a year.

Interrogative usages of 'who,' 'whom,' 'whose,' and 'which' can be distinguished from relative pronoun usages because interrogative adjectives are not preceded by a noun phrase or clause. Also, an input sentence which contains an interrogative adjective is followed by a question mark.

The usage of 'that' as a demonstrative pronoun is either not preceded by a clause or noun phrase, or not followed by a verb phrase or a clause introduced by a definite noun phrase. As a demonstrative pronoun, 'that' occurs in sentence structures such as (6) and (7).

6) That is John.

7) I gave you that because I thought you needed it.

As a demonstrative adjective 'that' is followed by a singular noun phrase which is not definite.

8) That man is John.

The conjunctive usage of 'that' introducing a noun clause in a cleft sentence can be detected according to rules in the grammar. Occurrences of conjunctive 'that' other than in cleft sentences can be distinguished from relative 'that.' In many cases conjunctive 'that' is not preceded by a noun phrase or clause, as in (9) and (10). In some cases conjunctive 'that' is preceded by the adverb 'so' modifying an adverb or adjective, as in (11), or the 'that' clause is preceded by an adverbial phrase like 'in such a manner,' as in (12).

9) That John was sick worried Mary.

10) A fourth characteristic of a compound is that the number and relationship of the atoms composing it must always be the same.

11) He ran so fast that I lost him.

12) A car should be driven in such a manner that it will last for a long time.

Conjunctive 'that' introducing a noun clause which is in apposition to the noun phrase which precedes 'that' often meets the environmental specifications for a relative pronoun

in the grammar. Notice in (13) 'that' is preceded by a noun phrase and followed by a clause with a definite noun phrase subject. However, no rules in the grammar can combine the occurrence of 'that' with the supposed relative clause, because the clause is complete. In cases like these, the rules in the grammar specify that the noun phrase preceding 'that' be examined to see if its head is one of a class of nouns which can be appositively modified by a 'that' clause.

13) The fact that microscopes had not reached a satisfactory stage in development...

The rules in the grammar which allow for the combination of a relative pronoun with the rest of the relative clause that follows it in an input sentence, specify that a copy of the relative pronoun and the constituent formed for the relative clause be put on the list of anaphora to be resolved when processing of the sentence is completed. The relative clause constituent is then combined with either the noun phrase constituent which directly precedes it in the input sentence or the clause constituent which precedes it. When this combination is made, the rules specify that the relation RELCL pointing to the relative clause constituent be put on the definitional list of the noun phrase or clause constituent and the relation RELCL-1 pointing to the noun phrase or clause constituent be put on the definitional list of the relative clause constituent.

IIId. Detection of Definite Noun Phrases

In most cases, a definite noun phrase composed of the definite determiner 'the' followed by a noun which is modified is not anaphoric. Only those noun phrase constituents formed for occurrences of the definite determiner 'the' followed by a noun which is not modified are put on the list of anaphora to be resolved.

IIIe. Preparation for Resolution of Anaphora

After translation of an input sentence to the internal information structures is completed, each item from the list of anaphora formed for the sentence is examined to see what type of anaphora it is. If the item is an instance of pronominal anaphora, the function PRONOUNS is called with the pronoun and the constituent in which it occurs as arguments. If the item is a noun phrase modified by a demonstrative adjective, the function DEMON is called with the noun phrase constituent as its argument. If the item is a relative pronoun, the function RELPRONOUNS is called with the relative pronoun and the relative clause constituent as its arguments. Finally, if the item is a definite noun phrase, the function LOOKTOPICS is called with the noun phrase constituent as its argument.

CHAPTER IV: THIRD PERSON PRONOUNS

The heuristics for finding the antecedent of a third person pronoun in a discourse were derived from a study of some relationships between a pronoun and its antecedent which occur frequently in discourse and from rules of English pronominalization. Different heuristics will be applied to find the antecedent of a third person pronoun depending on the pronoun's relation with a verbal constituent. The first step in deriving the antecedent of a third person pronoun is to determine what the pronoun's relation is with a verbal constituent.

The pronoun whose antecedent is sought is bound to the variable ANAPH in the function PRONOUNS given below. The constituent in which this pronoun occurs is bound to the variable CONSTIT. If CONSTIT is a verbal constituent, and ANAPH is in a subject relation with it, then the function SUBJPRO is called to derive the antecedent of ANAPH. If CONSTIT is a verbal constituent, and ANAPH is in an object relation with it, then the function OBJPRO is called. If ANAPH is a possessive modifier of the subject of a verbal constituent, then POSSSUBJ is called to derive the antecedent of ANAPH. If ANAPH is a possessive modifier of the object of a verbal constituent, then POSSOBJ is called. Finally, if ANAPH is in none of the above relations with a verbal constituent, then

the function DEPSEARCH is called to derive its antecedent. The function PRONOUNS examines ANAPH and CONSTIT to determine which of these situations exists.

```
(PRONOUNS  (LAMBDA  (ANAPH CONSTIT)
     (COND
     ((EQ (GETSUBJ CONSTIT) ANAPH)(SUBJPRO ANAPH CONSTIT))
     ((EQ (GETOBJ CONSTIT) ANAPH)(OBJPRO ANAPH CONSTIT))
     ((PSUBJMOD ANAPH CONSTIT)(POSSSUBJ ANAPH CONSTIT))
     ((POBJMOD ANAPH CONSTIT)(POSSOBJ ANAPH CONSTIT))
     (T (DEPSEARCH ANAPH CONSTIT))                    )))
```

IVa. Pronouns as Subjects of Verbal Constituents

The analysis of patterns of anaphora has revealed that, if a pronoun is the subject of a verbal constituent formed in processing a discourse, in most cases the antecedent of the pronoun is the subject of the preceding verbal constituent. However, if the pronoun is the subject of a verbal constituent which was derived from a preposed adverbial subordinate clause, the subject of the verbal constituent on which the subordinate clause constituent is dependent must be examined first as an antecedent candidate before the subjects of preceding verbal constituents are examined. If the antecedent of the pronominal subject of a verbal constituent is not the subject of the preceding verbal constituent, then the subject of the next preceding verbal constituent should be examined. If the subjects

of the preceding verbal constituents are not the antecedent of the pronoun, then the main topics list is searched for the antecedent. These procedures for deriving the antecedent of a pronoun in a subject relation with a verbal constituent are embodied in the function SUBJPRO given and explained below.

```
(SUBJPRO  (LAMBDA (ANAPH CONSTIT)
     (PROG(CANDIDATE SWITCH SWITCH1)
     (COND
     ((PSUBORD CONSTIT)(AND (SETQ CANDIDATE (GETSUBJ(GETNXTVB
                                                  CONSTIT)))
                            (SETQ SWITCH T)
                            (GO L2))))
L1   (SETQ CANDIDATE (GETSUBJ (GETPREVB CONSTIT)))
     (SETQ SWITCH NIL)
L2   (COND
     ((AND(SYNTCHK CANDIDATE ANAPH)(MEMBMTL CANDIDATE))
       (RETURN (REPLACE ANAPH CONSTIT CANDIDATE))))
     (COND (SWITCH (GO L1)))
     (COND (SWITCH1 (GO L3)))
     (SETQ CANDIDATE (GETSUBJ(GETPREVB(GETPREVB CONSTIT))))
     (SETQ SWITCH1 T)
     (GO L2)
L3   (RETURN (REPLACE ANAPH CONSTIT (LOOKUP ANAPH CONSTIT)))  )))
```

The first step in deriving the antecedent of ANAPH, the pronominal subject of the verbal constituent CONSTIT, is to see if CONSTIT was derived from a preposed adverbial subordinate clause. The first conditional statement in SUBJPRO makes such a check. If (PSUBORD CONSTIT) returns true, then:

i) The antecedent candidate is the subject of the verbal constituent on which CONSTIT is dependent.

ii) SWITCH is set true, so if this antecedent candidate is not the correct antecedent, a jump can be made back to L1 where the subject of the preceding verbal constituent is considered as an antecedent candidate.

iii) A jump is made to L2 where CANDIDATE is examined.

An antecedent candidate of a third person pronoun must be examined syntactically to see if it agrees in gender and number with the pronoun. This syntactic examination is accomplished in the statement (SYNTCHK CANDIDATE ANAPH). A semantic test must also be made of CANDIDATE since in some cases an antecedent candidate may agree with the pronoun in gender and number, but not be the correct antecedent of the pronoun. In (14), the antecedent candidate 'a second feature of the plasma membrane' agrees with the pronoun 'it' in gender and number, but is not the correct antecedent.

14) A second feature of <u>the plasma membrane</u> is the method of its formation.
<u>It</u> behaves as though it were a product of chemical reactions.

If a pronoun occurs as the subject of a verbal constituent formed for a discourse, this indicates that what the pronoun refers to is a current topic of the discourse. A reasonable semantic test to make of an antecedent candidate of a pronominal subject is to see if it is a prominent member of the list of current main topics of the discourse. This semantic check is accomplished by (MEMBMTL CANDIDATE).

If CANDIDATE meets the syntactic and semantic criteria, then ANAPH is replaced in CONSTIT by a pointer to CANDIDATE. If CANDIDATE does not meet the syntactic and semantic criteria, then (since SWITCH = T) a jump is made to L1 where CANDIDATE is set to the subject of the preceding verbal constituent. CANDIDATE is again examined syntactically and semantically.

If CONSTIT was not derived from a preposed adverbial subordinate clause, the subject of the preceding verbal constituent should be considered as an antecedent candidate. If this is the case, (PSUBORD CONSTIT) will return NIL and SUBJPRO will fall through to L1. The candidate derived at L1 in SUBJPRO is examined. If CANDIDATE does not meet the criteria, then the subject of the next preceding verbal constituent should be considered. CANDIDATE is set to the subject of the next preceding verbal constituent in the statement (SETQ CANDIDATE (GETSUBJ (GETPREVB (GETPREVB CONSTIT)))).

In those cases in which the subjects of the preceding verbal constituents do not agree with the pronoun in gender and number, the antecedent may be found by searching the list

of current main topics of the discourse until an item is found that agrees with the pronoun in gender and number and combines acceptably with the conceptual context of the pronoun. The primitive LOOKUP searches the main topics list for such an item and returns it as its value.

Consider the following example. The input sentence and the translated structures are given. Figure 3 gives a trace of the functions called in carrying out the procedures described above.

INPUT SENTENCE: If it grows larger, the cell will die because it cannot absorb enough of the necessary materials.

G1 = (TOK grow1)(MOD larger1)(SUBJ it1)(IF-1 G2)

G2 = (TOK die1)(SUBJ G3)(SUBJ G3)(IF G1)(BECAUSE G4)

G3 = (TOK cell1)(DET the1)(SUBJ-1 G2)

G4 = (TOK absorb1)(MOD cannot1)(SUBJ it1)(OBJ G5)(BECAUSE-1 G2)

G5 = (TOK enough1)(OBJ-1 G4)(OF G6)

G6 = (TOK material1)(DET the1)(MOD necessary2)(OF-1 G5)

FUNCTION CALLED	ARGUMENTS	VALUE	VARIABLES
PRONOUNS	it1 G1	T	
SUBJPRO	it1 G1	T	
PSUBORD	G1	T	

CANDIDATE = G3

SYNTCHK	G3 it1	T
MEMBMTL	G3	T
REPLACE	it1 G1 G3	T
PRONOUNS	it1 G4	T
SUBJPRO	it1 G4	T
PSUBORD	G4	NIL

CANDIDATE = G3

SYNTCHK	G3 it1	T
MEMBMTL	G3	T
REPLACE	it1 G4 G3	T

Figure 3. Trace of procedure for deriving antecedent of pronominal subject.

IVb. Pronouns as Objects of Verbal Constituents

If a third person pronoun is the object of a verbal constituent, modifiers of the subject of that constituent often contain the antecedent of the pronoun. There are some special constructions in which the antecedent of a pronominal object cannot occur in the subject constituent. If the antecedent is not found in the subject constituent of the verbal constituent, another candidate to consider which is often the antecedent is the object of the preceding verbal constituent. The procedures for deriving the antecedent of

a pronominal object are contained in the function OBJPRO given below.

```
(OBJPRO  (LAMBDA  (ANAPH CONSTIT)
     (PROG(CANDIDATE SUBJECT)
     (COND
     ((PSUBORD CONSTIT)(RETURN(DEPSEARCH ANAPH CONSTIT))))
     (SETQ SUBJECT (GETSUBJ CONSTIT))
     (COND
     ((PSENTMOD SUBJECT)(GO L4))
     ((AND (PPART SUBJECT)(NOT(PPOSSMOD SUBJECT)))(GO L4)))
     (COND
     ((SETQ CANDIDATE (LOOKMOD SUBJECT ANAPH CONSTIT))
       (RETURN(REPLACE ANAPH CONSTIT CANDIDATE))))
L4   (SETQ CANDIDATE (GETOBJ(GETPREVB CONSTIT)))
     (COND
     ((AND (SYNTCHK CANDIDATE ANAPH)(CONCEPTCHK CANDIDATE
       ANAPH CONSTIT))
       (RETURN (REPLACE ANAPH CONSTIT CANDIDATE))))    )))
```

If CONSTIT is a verbal constituent which was derived from a preposed adverbial subordinate clause, the antecedent of ANAPH will be found by the procedures specified in the function DEPSEARCH. The first conditional of the function OBJPRO handles this.

Two constructions have been detected in which the antecedent of a pronominal object, ANAPH, cannot occur in the subject constituent, SUBJECT, because the constructions allow only backward pronominalization. If SUBJECT is composed of a noun modified by a 'that' clause which is in apposition to the noun, the antecedent of ANAPH cannot occur in SUBJECT, but must occur in some preceding portion of the discourse. If SUBJECT is a participal which is not possessively modified, then the antecedent of ANAPH cannot occur in SUBJECT. In either case, the object of the preceding verbal constituent is considered as an antecedent candidate of ANAPH. The second conditional of OBJPRO examines SUBJECT to see if it is formed in either of the above ways, and if so a jump is made to L4. (15 and (16) are surface string examples of these constructions.

15) The knowledge that John was sick worried him.

16) Realizing that John was sick worried him.

If SUBJECT is not one of these two constructions, the noun or noun phrase modifiers of the head of SUBJECT are examined to see if one agrees with ANAPH in gender and number and combines acceptably with the conceptual context of ANAPH. The function LOOKMOD examines the modifiers and, if one is found that meets the syntactic and semantic criteria, LOOKMOD returns it as its value. CANDIDATE is set to the value of LOOKMOD, and ANAPH is replaced in CONSTIT by CANDIDATE.

If LOOKMOD does not find a modifier of SUBJECT which

meets the criteria, it returns NIL as its value, and the value of (SETQ CANDIDATE (LOOKMOD SUBJECT ANAPH)) is NIL. The function OBJPRO then falls through to L4.

If SUBJECT is one of the special constructions described above, or if no antecedent candidate is found in the modifiers of SUBJECT which meet the syntactic and semantic criteria, then the object of the preceding verbal constituent is examined as an antecedent candidate. CANDIDATE is set to the object of the preceding verbal constituent at L4 in OBJPRO. CANDIDATE must be checked to see if it agrees with ANAPH in gender and number, and to see if it combines acceptably with the conceptual context of ANAPH.

Following are two examples of sentences with pronominal objects. In the first example the antecedent of the pronominal object is the object of the preceding verbal constituent. A trace of the functions called in deriving the antecedent of the pronominal object in the first example is given in Figure 4. In the second example the antecedent of the pronominal object is in the relative clause modifying the subject.

INPUT SENTENCE: Robert Hooke cut very thin sections of cork and examined them under the microscope.

G1 = (TOK cut1)(SUBJ Robert Hooke)(OBJ G2)(EL1-1 G6)

G2 = (TOK section2)(MOD G3)(OF cork1)

G3 = (TOK thin1)(MOD very1)(MOD-1 G2)

G4 = (TOK examine1)(SUBJ Robert Hooke)(OBJ them1)(UNDER G5)
(EL2-1 G6)

G5 = (TOK microscope2)(DET the1)(UNDER-1 G4)

G6 = (TOK and1)(EL1 G1)(EL2 G4)

FUNCTION CALLED	ARGUMENTS	VALUE	VARIABLES
PRONOUNS	them1 G4	T	
OBJPRO	them1 G4	T	
PSUBORD	G4	NIL	SUBJECT=Robert Hooke
PSENTMOD	Robert Hooke	NIL	
PPART	Robert Hooke	NIL	
PPOSSMOD	Robert Hooke	NIL	
LOOKMOD	R. Hooke them1	NIL	
			CANDIDATE=NIL
GETPREVB	G4	G1	
GETOBJ	G1	G2	
			CANDIDATE=G2
SYNTCHK	G2 them1	T	
CONCEPTCHK	G2 G4	T	
REPLACE	them1 G4 G2	T	

Figure 4. Trace of procedure for deriving antecedent of pronominal object.

INPUT SENTENCE: Students who Dr. Smith has taught don't like him.

G10 = (TOK student2)(OBJ-1 G11)(SUBJ-1 G12)

G11 = (TOK teach1)(OBJ who G10)(SUBJ Dr. Smith)

G12 = (TOK like1)(OBJ him1)(SUBJ G1)(NEG not2)

IVc. Possessive Pronouns Modifying the Subject of a Verbal Constituent

When a possessive third person pronoun modifies the subject of a verbal constituent, procedures similar to those for third person pronouns which are subjects are applied in order to determine the antecedent of the possessive pronoun. However, certain constructions in which a possessive pronoun can occur indicate special procedures which are often successful in finding the antecedent of the pronoun. The subject constituent, CONSTIT, in which the possessive pronoun, ANAPH, occurs is examined to see if it constitutes one of these special constructions, and if so the procedures indicated are followed. If the object of the verbal constituent of which CONSTIT is the subject is a 'that' clause, it should be examined for the antecedent of ANAPH. If CONSTIT is not one of the special constructions and the object constituent is not a 'that' clause, then the subject of the preceding verbal constituent is examined to see if it is possessively modified. If it is, this possessive modifier is usually the antecedent of ANAPH. If the antecedent of the possessive pronoun is not found by any of these procedures, SUBJPRO is called with ANAPH and the verbal constituent of which CONSTIT is subject as arguments. SUBJPRO considers the subjects of preceding verbal constituents as antecedent candidates, and if no antecedent is found it searches the main topics list for the

antecedent.

```
(POSSSUBJ  (LAMBDA (ANAPH CONSTIT) (PROG(CANDIDATE THAT)
     (COND
     ((SETQ CANDIDATE (LOOKLEX CONSTIT (GETVERB CONSTIT)))
       (RETURN (REPLACE ANAPH CONSTIT CANDIDATE)))
     ((AND (PCONJ2 CONSTIT)(SYNTCHK (OTHEREL CONSTIT) ANAPH))
       (RETURN (REPLACE ANAPH CONSTIT (OTHEREL CONSTIT))))
     ((AND (PTHAT (SETQ THAT (GETOBJ(GETVERB CONSTIT))))
           (SETQ CANDIDATE (LOOKMOD THAT ANAPH CONSTIT)))
       (RETURN (REPLACE ANAPH CONSTIT CANDIDATE)))
     ((SETQ CANDIDATE (LOOKMOD (GETSUBJ(GETPREVB(GETVERB CONSTIT)))
                                    ANAPH CONSTIT))
       (RETURN (REPLACE ANAPH CONSTIT CANDIDATE))))
     (RETURN (SUBJPRO ANAPH (GETVERB CONSTIT)))    )))
```

If a noun modified by a possessive pronoun has occurred recently in the discourse with possessive modification, this possessive modifier will usually be the antecedent of the possessive pronoun, as in (17).

17) Harvey had to imagine the presence of <u>the smallest blood vessels, the capillaries</u>.
    In 1632, <u>their</u> presence...

LOOKLEX sees if the head noun of CONSTIT occurs in a constituent formed for a closely preceding portion of the

discourse. It sees if the preceding occurrence of the head noun has possessive modification, and if so it checks for syntactic agreement of this possessive modifier and ANAPH. If the syntactic criteria are met by this possessive modifier, LOOKLEX returns it as its value, and ANAPH is replaced by it in CONSTIT. If the criteria are not met, or if no occurrence of the head noun of CONSTIT is found with posssessive modification, LOOKLEX returns NIL, and the conditional statement in POSSSUBJ falls through to the next pair.

In the second pair of the conditional statement, CONSTIT is examined to see if it is the second element of a conjoined form. If two noun phrases are conjoined and a possessive pronoun modifies the second noun phrase of the conjoined form, the antecedent of the possessive pronoun is usually the first noun phrase of the conjoined form. Consider (18):

18) ...<u>living things</u> and <u>their</u> environments...

PCONJ2 returns true if CONSTIT is the second element of a conjoined form. SYNTCHK checks for syntactic agreement of the other element of the conjoined form and ANAPH. If both PCONJ2 and SYNTCHK return true, ANAPH is replaced by the other element of the conjoined form.

The third pair of the conditional statement in POSSSUBJ examines the object of the verbal constituent of which CONSTIT is subject. If the object of the verbal

constituent is a 'that' clause, it is examined for the antecedent of a possessive pronoun modifying the subject. Consider (19):

19) Her boyfriends think that Mary is beautiful.

If the object is a 'that' clause, PTHAT returns true, and LOOKMOD examines the elements of THAT for the antecedent of ANAPH. If it is found, ANAPH is replaced by it.

The fourth part of the conditional examines the subject of the verbal constituent preceding the verbal constituent of which CONSTIT is the subject. LOOKMOD looks at the modifiers of the preceding subject constituent to see if one agrees with ANAPH syntactically and conceptually, as in (20).

20) Harvey's methods were carefully developed and his thinking was clear.

Finally, if the antecedent of ANAPH has not been found, SUBJPRO is called. In SUBJPRO, the subjects of preceding verbal constituents and elements of the main topics list are considered as antecedent candidates.

IVd. Possessive Pronouns Modifying the Object of a Verbal Constituent

If a possessive pronoun modifies the object of a verbal constituent, its antecedent is often the subject of that verbal constituent, or one of the elements of a phrase

or clause modifier of the subject. In some cases the antecedent of the possessive pronoun is the object of the preceding verbal constituent. The function POSSOBJ handles the case in which the antecedent of the possessive pronoun is the entire subject constituent, and then it calls OBJPRO to examine the modifiers of the subject constituent and the object of the preceding verbal constituent.

```
(POSSOBJ  (LAMBDA (ANAPH CONSTIT)(PROG(CANDIDATE)
    (COND
    ((AND(SYNTCHK(SETQ CANDIDATE (GETSUBJ(GETVERB CONSTIT)))
                ANAPH)
        (CONCEPTCHK CANDIDATE ANAPH CONSTIT))
      (RETURN (REPLACE ANAPH CONSTIT CANDIDATE))))
    (RETURN (OBJPRO ANAPH (GETVERB CONSTIT)))   )))
```

In (21) a possessive pronoun, 'its,' modifies the subject of a verbal constituent and the object of the same verbal constituent. The antecedent of the occurrence of 'its' modifying the subject is found to be the subject of the preceding verbal constituent, a constituent formed for the surface string 'a newly formed cubical cell,' by the heuristics given in IVc. The antecedent of the occurrence of 'its' modifying the object is the modifier of the subject, the first occurrence of 'its' which has already been associated with its antecedent and carries all of the syntactic and

definitional properties of this antecedent.

21) If a newly formed cubical cell measures 10/2500 inch, its growth may double its dimension.

In (22) the antecedent of the possessive pronoun modifying the object is the entire subject constituent.

22) The tiny bones of the ear have reached their maximum size before birth.

IVe. Other Occurrences of Third Person Pronouns

The procedure for deriving the antecedents of all occurrences of third person pronouns in the translated structures which are not subjects or objects, or possessive modifiers of subjects or objects, of main verbal constituents involves traversing the dependencies made explicit in the translated structures by relational links between constituents. The function DEPSEARCH traverses the dependency structure, starting at the constituent in which the anaphoric pronoun occurs, examining all noun phrase items for syntactic and conceptual agreement with the pronoun. DEPSEARCH traverses the reverse links between constituents until it reaches the verbal constituent. Then it follows the forward relational links of all of the other constituents which are dependent on the verbal constituent.

Because of the transitive dependency of constituents on a verbal constituent and the method of searching up to the verbal constituent and down through all of the other

constituents which are dependent on that verbal constituent, most instances of anaphoric pronouns in translated structures which were derived from input sentences in which backward pronominalization has occurred can be resolved. In the following input sentence backward pronominalization has occurred. The pronoun 'he' occurs in G2 which has the reverse link THAT-1 pointing to G1. G1 has the reverse link SUBJ-1 pointing to G3, which is a verbal constituent. DEPSEARCH examines all other constituents which are dependent on the verb of G3, and finds the antecedent, John, in an OBJ relation with the verb of G3.

INPUT SENTENCE: Realizing that he was sick worried John.

G1 = (TOK realizing2)(THAT G2)(SUBJ-1 G3)

G2 = (TOK was1)(SUBJ he1)(PR-ADJ sick2)(THAT-1 G1)

G3 = (TOK worry2)(SUBJ G1)(OBJ John1)

CHAPTER V: DEMONSTRATIVE ADJECTIVES MODIFYING NOUN PHRASES

Several relationships have been detected which occur frequently between an anaphoric noun phrase introduced by a demonstrative adjective and its antecedent. The antecedent of an anaphoric noun phrase introduced by a demonstrative adjective may be a noun phrase constituent whose head noun is lexically identical to the head noun of the anaphoric expression.

23a) In 1628 he published a small book.

23b) This book had only 72 pages.

Sometimes the head noun of the antecedent noun phrase constituent is synonymous with the head noun of the anaphoric expression.

24a) In 1628 he published a small book.

24b) This small report is one of the finest examples of the scientific method.

24c) The importance of this work was...

The antecedent may be in a SUP relation with the head noun of the anaphoric constituent, i.e. the antecedent may be a particular member of the more general set specified by the anaphoric noun phrase.

25) Embryology deals with the early development of the individual, and recent experiments in this field...

When the head noun of the anaphoric noun phrase constituent relates to a verbal form, and that verbal form heads the preceding verbal constituent, the antecedent of the anaphoric noun phrase is the entire preceding verbal constituent.

26) Tiny cells move in a peculiar manner.
This movement can be viewed through a microscope.

The procedure for deriving the antecedent of an anaphoric noun phrase introduced by a demonstrative adjective involves determining if any of the above relationships exist between the anaphoric constituent and a preceding discourse constituent. The function DEMON makes calls to the primitives LOOKLEX1, LOOKSYN, LOOKSUP, and LOOKVERB, which examine the anaphoric constituent ANAPH and preceding discourse elements to find an antecedent which is in one of these four relationships with ANAPH.

```
(DEMON  (LAMBDA  (ANAPH) (PROG (CANDIDATE)
    (COND
    ((SETQ CANDIDATE (LOOKLEX1 ANAPH (GETVERB ANAPH)))
      (RETURN (REPLACE1 ANAPH CANDIDATE)))
    ((SETQ CANDIDATE (LOOKSYN ANAPH))
      (RETURN (ADDLINK ANAPH CANDIDATE)))
    ((SETQ CANDIDATE (LOOKSUP ANAPH))
```

```
    (RETURN (ADDLINK ANAPH CANDIDATE)))
((SETQ CANDIDATE (LOOKVERB ANAPH (GETVERB ANAPH)))
    (RETURN (ADDLINK ANAPH CANDIDATE)))    ))))
```

Notice that if LOOKLEX finds an element in the preceding constituents for the discourse which is lexically identical to the head noun of ANAPH, ANAPH is replaced in all of the constituents in which it occurs by this element. However, if the antecedent is in one of the other relations with ANAPH, the relational link ANAPHORIC pointing to the antecedent constituent is added to the constituent ANAPH and the relational link ANAPHORIC-1 pointing to ANAPH is added to the antecedent constituent.

The translated structures for (23a), (23b), and (24b) follow. G4 and G6 are anaphoric noun phrase constituents. The head noun of G2 is lexically identical with the head noun of G4, so G4 is deleted after resolution of anaphora and the pointer to G4 in G3 is replaced by a pointer to G2. The head noun of G2 is in a synonym relation with the head noun of G6, so the pair (ANAPHORIC G2) is added to G6 and the pair (ANAPHORIC-1 G6) is added to G2.

BEFORE RESOLUTION OF ANAPHORA:

G1 = (TOK publish1)(SUBJ he1)(IN 1628)(OBJ G2)

G2 = (TOK book2)(DET A1)(MOD small2)(OBJ-1 G1)

G3 = (TOK have1)(SUBJ G4)(OBJ G5)

G4 = (TOK book2)(DEMADJ this1)(SUBJ-1 G3)

G5 = (TOK page1)(QUANT only1 72)(OBJ-1 G3)

G6 = (TOK report1)(DEMADJ this1)(MOD small1)(EL1-1xG7)

G7 = (TOK is1)(EL1 G6)(EL2 G8)

G8 = (TOK one1)(SUP G9)

G9 = (TOK example2)(DET the1)(MOD finest1)(SUP-1 G8)(OF G10)

G10= (TOK method2)(DET the1)(MOD scientific1)(OF-1 G9)

AFTER RESOLUTION OF ANAPHORA:

G1 = (TOK publish1)(SUBJ he1)(IN 1628)(OBJ G2)

G2 = (TOK book2)(DET a1)(MOD small2)(OBJ-1 G1)(ANAPHORIC-1 G6)
(SUBJ-1 G3)

G3 = (TOK have1)(SUBJ G2)(OBJ G5)

G5 = (TOK page1)(QUANT only1 72)(OBJ-1 G3)

G6 = (TOK report1)(DEMADJ this1)(MOD small1)(EL1-1 G7)(ANAPHORIC G2)

G7 = (TOK is1)(EL1 G6)(EL2 G8)

G8 = (TOK one1)(SUP G9)

G9 = (TOK example2)(DET the1)(MOD finest1)(SUP-1 G8)(OF G10)

G10= (TOK method2)(DET the1)(MOD scientific1)(OF-1 G9)

CHAPTER VI: RELATIVE PRONOUNS

During processing of a sentence which contains a relative clause, the constituent formed for the relative clause is made dependent by the relational link RELCL-1 on the noun phrase constituent or the verbal constituent formed for the smallest noun phrase or the clause which precedes it in the input sentence. The first step in the procedure for deriving the antecedent of a relative pronoun is to get this constituent on which the relative clause constituent is dependent and examine it to see if it is a noun phrase constituent or a verbal constituent. CANDIDATE is set to the constituent on which the relative clause constituent is dependent in the first statement of the function RELPRONOUNS. CANDIDATE is examined to see if it is a noun phrase constituent or a verbal constituent in the first conditional statement.

```
(RELPRONOUNS  (LAMBDA  (ANAPH CONSTIT) (PROG(CANDIDATE)
    (SETQ CANDIDATE ( GETREL CONSTIT))
    (COND
    ((PNP CANDIDATE)(GO L5))
    ((SYNTCHK1  (GETSUBJ CANDIDATE) ANAPH CONSTIT)
      (RETURN (ADDTO ANAPH CONSTIT (GETSUBJ CANDIDATE)))))
L5  (COND
```

```
((AND(PSING CANDIDATE)(PNODET CANDIDATE))
  (COND
  ((SYNTCHK1 (GETPREP CANDIDATE) ANAPH CONSTIT)
    (RETURN (ADDTO ANAPH CONSTIT (GETPREP CANDIDATE))))
  ((SYNTCHK1 CANDIDATE ANAPH CONSTIT)
    (RETURN (ADDTO ANAPH CONSTIT CANDIDATE)))))
((SYNTCHK1 CANDIDATE ANAPH CONSTIT)
  (RETURN (ADDTO ANAPH CONSTIT CANDIDATE)))
((SYNTCHK1 (GETPREP CANDIDATE) ANAPH CONSTIT)
  (RETURN (ADDTO ANAPH CONSTIT (GETPREP CANDIDATE))))
((SYNTCHK1  (GETCONJ CANDIDATE ANAPH CONSTIT)
  (RETURN (ADDTO ANAPH CONSTIT (GETCONJ CANDIDATE))))
((SYNTCHK1 (GETSUBJ(GETVERB CANDIDATE)) ANAPN CONSTIT)
  (RETURN (ADDTO ANAPH CONSTIT (GETSUBJ(GETVERB CANDIDATE))))))
                                                  )))
```

If CANDIDATE is a noun phrase constituent, a jump is made to L5 where it is examined further. If CANDIDATE is a verbal constituent, this indicates that the relative clause constituent, CONSTIT, was an extraposed relative clause in the input sentence. The subject of the verbal constituent CANDIDATE is the antecedent of the relative pronoun, ANAPH, in most cases. The subject is examined by SYNTCHK1 to see if it agrees in gender with ANAPH, and to see if it agrees in number with the verb of CONSTIT if ANAPH is in a SUBJ relation with the verb of CONSTIT or if the verb of CONSTIT is a 'to be' verb and

ANAPH is in an EL1 relation with it.

If the subject of the verbal constituent CANDIDATE meets these syntactic specifications, it is added to CONSTIT in the same relation to the verb of CONSTIT as ANAPH. The inverse relation is put on the subject constituent, and the relations RELCL and RELCL-1 are deleted from CANDIDATE and CONSTIT. Consider the following example.

INPUT SENTENCE: A man died who had been sick for a year.

TRANSLATED STRUCTURES BEFORE RESOLUTION OF ANAPHORA:

G1 = (TOK man1)(DET a1)(SUBJ-1 G2)

G2 = (TOK die2)(SUBJ G1)(RELCL G3)

G3 = (TOK sick2)(SUBJ who2)(FOR G4)(RELCL-1 G2)

G4 = (TOK year1)(DET a1)(FOR-1 G3)

ANAPH = who2 CONSTIT = G3

FUNCTION CALLED	ARGUMENTS	VALUE	VARIABLES
RELPRONOUNS	who2 G3	T	
			CANDIDATE = G2
PNP	G2	NIL	
SYNTCHK1	G1 who2 G3	T	
ADDTO	who2 G3 G1	T	

Figure 5. Trace of procedure for deriving antecedent of relative pronoun.

TRANSLATED STRUCTURES AFTER RESOLUTION OF ANAPHORA:

G1 = (TOK man1)(DET a1)(SUBJ-1 G2 G3)

G2 = (TOK die2)(SUBJ G1)

G3 = (TOK sick2)(SUBJ who2 G1)(FOR G4)

G4 = (TOK year1)(DET a1)(FOR-1 G3)

At L5 in RELPRONOUNS the noun phrase constituent CANDIDATE is examined. If CANDIDATE is singular and has no determiner, then a check is made to see if there are any reverse prepositional links in CANDIDATE indicating that CANDIDATE is the noun phrase of a prepositional phrase modifying another noun phrase. If this is the case, the larger noun phrase is usually the antecedent of ANAPH. Consider (27), a surface string example in which the noun phrase directly preceding 'that' is singular and has no determiner, but is part of a prepositional phrase modifying another noun phrase. The larger noun phrase is the antecedent.

27) Man has a type of body symmetry that is known as bilateral.

If CANDIDATE, a singular noun phrase with no determiner, is not prepositionally dependent on another noun phrase, (GETPREP CANDIDATE) returns NIL and SYNTCHK1 returns NIL. If CANDIDATE is dependent on another noun phrase, but this noun phrase constituent, returned as the value of (GETPREP CANDIDATE), does not agree syntactically with ANAPH, then SYNTCHK1 returns NIL. In either case the embedded conditional falls through to

the second pair. In the second pair, CANDIDATE is examined for syntactic agreement with ANAPH.

It has been found that, in most cases, if CANDIDATE is not a singular noun phrase constituent with no determiner, if it agrees syntactically with ANAPH, it is the antecedent of ANAPH even if it is part of a larger noun phrase. A check for syntactic agreement of CANDIDATE and ANAPH is made in the second pair of the conditional which begins at L5 in RELPRONOUNS.

If CANDIDATE, the smallest noun phrase preceding the relative clause in the input sentence, does not agree syntactically with ANAPH, then CANDIDATE is examined by the function GETPREP to see if it is prepositionally dependent on another noun phrase constituent. If it is, this noun phrase constituent is examined for syntactic agreement with ANAPH. Consider (28), a surface string example in which the smallest noun phrase preceding the relative clause does not agree syntactically with the relative pronoun, but is prepositionally dependent on a noun phrase which does agree syntactically with the relative pronoun.

28) Buildings in the city which are old are being torn down.

If CANDIDATE is not prepositionally dependent on another noun phrase constituent, it may be part of a conjoined noun phrase. The entire conjoined noun phrase constituent should be the antecedent of ANAPH. CANDIDATE is examined in the call to GETCONJ in the fourth pair of the conditional to

see if it is an element of a conjoined form.

If these procedures are not successful in finding a noun phrase which agrees syntactically with ANAPH, CONSTIT must have been an extraposed relative clause in the input sentence which was conbined with the noun phrase constituent that preceded it rather than the verbal constituent for the clause that preceded it. Consider the translated structures before resolution of anaphora for the following input sentence.

INPUT SENTENCE: A man went to the fair who lost his mind.
TRANSLATED STRUCTURES BEFORE RESOLUTION OF ANAPHORA:
G1 = (TOK man1)(DET a1)(SUBJ-1 G2)
G2 = (TOK go1)(SUBJ G1)(TO G3)
G3 = (TOK fair1)(DET the1)(TO-1 G2)(RELCL G4)
G4 = (TOK lose1)(SUBJ who2)(OBJ G5)(RELCL-1 G3)
G5 = (TOK mind2)(POSS him1)(OBJ-1 G4)

The final pair of the conditional at L5 in RELPRONOUNS gets the verbal constituent on which CANDIDATE is dependent and examines the subject of this verbal constituent for syntactic agreement with ANAPH. This takes care of those cases when an extraposed relative clause is combined with the noun phrase preceding it rather than the clause preceding it.

TRANSLATED STRUCTURES AFTER RESOLUTION OF ANAPHORA:
G1 = (TOK man1)(DET a1)(SUBJ-1 G2 G4)

```
G2 = (TOK go1)(SUBJ G1)(TO G3)

G3 = (TOK fiar1)(DET the1)(TO-1 G2)

G4 = (TOK lose1)(SUBJ who2 G1)(OBJ G5)

G5 = (TOK mind2)(POSS him1)(OBJ-1 G4)
```

CHAPTER VII: DEFINITE NOUN PHRASES

An occurrence of an unmodified noun marked with a definite determiner in a discourse has an existing referent. The existence of this referent may be established in a preceding portion of the discourse, or the existence of the referent is assumed to be understood by the readers of the discourse. For example, an occurrence of the definite noun phrase 'the moon' in a discourse will not usually refer to a previous mention of 'moon' in the discourse. The existence of the referent is established by the common knowledge that 'Earth has a moon' shared by the readers of the discourse.

A discourse referent of an unmodified definite noun phrase may either be established explicitly or implicitly. An explicitly established discourse referent is an occurrence of the same word sense of the head noun of the definite noun phrase constituent in a noun phrase constituent formed for a preceding portion of the discourse. The existence of the referent may be implicitly established by a preceding discourse element which is in a relationship with the definite noun phrase which implies the existence of the referent. Consider (29), an example from Kartunnen (1968), which has both explicit and implicit discourse referents of definite noun phrases.

29) I was driving a car on the freeway the other day. Suddenly the engine began to make a funny noise. I stopped the car and when I opened the hood I saw that the radiator was boiling.

In (29), the referent of 'the car' is explicitly established. The referents of 'the engine,' 'the hood,' and 'the radiator' are implicitly established. The information that a car has as parts an engine, a hood, and a radiator, assumed to be understood by the reader, established the referents for the definite noun phrases 'the engine,' the hood,' and 'the radiator.'

The antecedent of a definite unmodified noun phrase is a noun phrase constituent formed for a preceding portion of the discourse whose head noun establishes, explicitly or implicitly, the referent of the definite noun phrase. The current list of main topics of the discourse is searched for a noun phrase constituent whose head noun 1) is lexically identical to the noun of the definite noun phrase constituent, or 2) has relational links in its lexical entry which establish a relation between it and the head noun of the definite noun phrase constituent. The function LOOKTOPICS makes such a search of the main topics list. In some cases, no antecedent will be found on the main topics list because the referent of the definite noun phrase is understood.

If a noun phrase constituent is found on the main

topics list whose head noun is lexically identical to the noun of the definite noun phrase constituent, the relation EQUIV pointing to this antecedent noun phrase constituent is added to the definitional list of the definite noun phrase constituent, and EQUIV pointing to the definite noun phrase constituent is added to the definitional list of the antecedent noun phrase constituent. If a noun phrase constituent is found on the main topics list whose head noun is in a relation, which is made explicit by relational links in the lexicon, with the noun of the definite noun phrase, then that relation is added to the definite noun phrase constituent and its inverse relation is added to the antecedent noun phrase constituent.

Consider the following example, in which the relational links HAS-PART and PART-OF are added to G1 and G5 during application of LOOKTOPICS, to make explicit the referent of the definite noun phrase 'the seed.'

INPUT SENTENCES: The coconut palm appears in newly formed tropical islands.

The seed is protected from the action of sea water.

BEFORE RESOLUTION OF ANAPHORA:

G1 = (TOK palm1)(ASSOC coconut2)(DET the1)(SUBJ-1 G2)

G2 = (TOK appear2)(SUBJ G1)(IN G3)

G3 = (TOK island2)(MOD tropical2 G4)(IN-1 G2)

G4 = (TOK formed1)(TIME newly2)(MOD-1 G3)

G5 = (TOK seed1)(DET the1)(SUBJ-1 G6)

G6 = (TOK protect2)(FROM G7)(SUBJ G5)

G7 = (TOK action1)(DET the1)(FROM-1 G6)(OF G8)

G8 = (TOK water1)(ASSOC sea2)(OF-1 G7)

AFTER RESOLUTION OF ANAPHORA:

G1 = (TOK palm1)(ASSOC coconut2)(DET the1)(SUBJ-1 G2)(HAS-PART G5)

G5 = (TOK seed1)(DET the1)(SUBJ-1 G6)(PART-OF G1)

CHAPTER VIII: CONCLUSION

The anaphora resolution subunit described is currently being programmed into the language processor of which the components described in Chapter II are parts. This language processor is the latest model of the Protosynthex system, which is being designed and programmed under the direction of Dr. Robert F. Simmons at the University of Texas at Austin.

The heuristics embodied in the LISP functions have been experimented with, by hand simulating a large number of sentences which contain the types of anaphora the heuristics handle. These sentences were of the style of English discourse used in elementary scientific articles, such as articles from high school science books and intermediate encyclopedias. For this limited style of English discourse, the heuristics showed a high degree of accuracy. The anaphoric expressions were detected in input sentences and associated with their antecedents, by following the heuristics stated in the LISP functions, in almost all cases.

APPENDIX

The primitives described below are used in the LISP functions which carry out the procedures for deriving the antecedents of certain types of anaphora.

Predicates:

The following predicates examine certain syntactic properties and definitional relations of constituents:

PCONJ2 returns true if the constituent which is its argument is the second element of a conjoined form.

PNODET returns true if the constituent which is its argument has no determiner.

PNP returns true if the constituent which is its argument is a noun phrase constituent.

POBJMOD is called with a third person pronoun and the constituent in which the pronoun occurs as arguments. It returns true if the constituent in which the pronoun occurs is in a OBJ relation with the verb of a verbal constituent, and the third person pronoun is a possessive modifier of this object.

PPART returns true if the constituent which is its argument has possessive modification.

PSENTMOD returns true if its argument is a noun phrase constituent which has a "that" clause in an appositive relation with it.

PSING returns true if its argument is singular.
PSUBJMOD is called with a third person pronoun and the constituent in which the pronoun occurs as arguments. It returns true if the constituent in which the pronoun occurs is in a SUBJ relation with the verb of a verbal constituent, and the pronoun is a possessive modifier of this subject.
PSUBORD returns true if the constituent which is its argument is a verbal constituent which was formed from a preposed adverbial subordinate clause.
PTHAT returns true if its argument is a "that" clause.

Retrieval Functions:

The following functions retrieve a constituent which is in a certain relation with the constituent that is the argument of the function.

GETCONJ is given a noun phrase constituent as its argument. If this noun phrase constituent is in an EL2 relation with a conjunction constituent, GETCONJ returns the conjoined noun phrase constituent as its value. Otherwise, it returns NIL.
GETNXTVB is given a verbal constituent which was formed from a preposed adverbial subordinate clause as its argument. It returns as its value the verbal constituent on which its argument is dependent.
GETOBJ checks to see if its argument is a member of the list of verbal constituents, and if so it returns the constituent

which is in an OBJ relation with this verbal constituent as its value. If its argument is not a verbal constituent, it returns NIL as its value.

GETPREP is given a noun phrase constituent as its argument. If this noun phrase constituent is prepositionally dependent on a noun phrase constituent, GETPREP returns the noun phrase constituent on which its argument is dependent as its value. If its argument is not prepositionally dependent on a noun phrase constituent, it returns NIL as its value.

GETPREVB is given a verbal constituent as its argument. It returns as its value the verbal constituent which precedes its argument in the list of verbal constituents formed for the discourse.

GETREL is given a relative clause constituent as its argument. It returns as its value the noun phrase or verbal constituent on which the relative clause constituent is dependent.

GETSUBJ checks to see if its argument is a member of the list of verbal constituents, and if so it returns the constituent which is in a SUBJ relation with this verbal constituent as its value. If its argument is not a verbal constituent, it returns NIL as its value.

GETVERB returns the verb on which the constituent which is its argument is dependent. The constituent may be directly dependent on a verbal constituent (i.e. in a SUBJ or OBJ relation with the verbal constituent) or transitively dependent on a verbal constituent.

OTHEREL is given an element of a conjoined form as its argument. It returns the other element of the conjoined form as its value.

Examination Predicates:

The following predicates examine syntactic and definitional properties of an anaphoric expression and an antecedent candidate to see if the antecedent candidate is the correct antecedent:

CONCEPTCHK is called with a third person pronoun, the constituent in which the pronoun occurs, and an antecedent candidate for the pronoun. If the pronoun is in a SUBJ or OBJ relation with a verbal constituent, CONCEPTCHK compiles a list of concepts which are SUP to the antecedent candidate and a list of concepts which are SUP to the verb of the verbal constituent by examining their lexical entries. CONCEPTCHK examines these lists to see if a concept SUP to the antecedent candidate and a concept SUP to the verb of the anaphoric pronoun will combine acceptably according to the list of concept combinations specified in the system. If the pronoun is in a SUBJ relation with the verb, concepts SUP to the antecedent candidate precede concepts SUP to the verb when they are compared with the list of acceptable concept combinations. If the pronoun is in an OBJ relation with the verb, concepts SUP to the antecedent candidate follow concepts SUP to the verb.

If the pronoun is a possessive modifier of a noun

phrase constituent. CONCEPTCHK examines the head noun that the possessive pronoun modifies to see if it relates to a verbal form (for example, the noun "growth" relates to a verbal form "grow"). If the noun does relate to a verbal form, CONCEPTCHK compiles lists of concepts SUP to the antecedent candidate and concepts SUP to the verbal form of the noun to see if any concepts SUP to the antecedent candidate combine acceptably with concepts SUP to the verbal form.

If the pronoun is a possessive modifier of a noun phrase constituent, but the head noun of the noun phrase constituent does not relate to a verbal form, then the lexical entry for the antecedent candidate of the pronoun is examined to see if any information indicates that the antecedent candidate can possess the noun phrase modified by the possessive pronoun.

MEMBMTL checks to see if its argument is a member of the current list of main topics for the discourse. If it is, MEMBMTL returns true, otherwise it returns NIL.

SYNTCHK is called with a third person pronoun and its antecedent candidate as its arguments. If the pronoun and the antecedent candidate agree in gender and number, SYNTCHK returns true, otherwise it returns NIL.

SYNTCHK1 is called with a relative pronoun, the relative clause constituent for that pronoun, and an antecedent candidate for the pronoun as its arguments. SYNTCHK1 checks for agreement in gender of the relative pronoun and the

antecedent candidate. If the relative pronoun is in a SUBJ relation with the verb of the relative clause constituent, or if the verb of the relative clause constituent is a 'to be' verb and the relative pronoun is in an EL1 relation with it, SYNTCHK1 also checks for agreement in number of the antecedent candidate and the verb of the relative clause constituent. If the antecedent candidate agrees syntactically with the relative pronoun and its relative clause constituent, SYNTCHK1 returns true, otherwise it returns NIL.

Search Functions:

The following functions apply certain search procedures to find an antecedent for an anaphoric expression:

DEPSEARCH is called with a third person pronoun and the constituent in which it occurs as its arguments. DEPSEARCH traverses the reverse relational links from the constituent, examining each noun phrase on which the constituent which contains the pronoun is dependent for syntactic and conceptual agreement with the pronoun If DEPSEARCH reaches the verbal constituent without finding a noun phrase constituent which agrees syntactically with the pronoun and fits into the conceptual context of the pronoun, it traverses the relational links of all other constituents which are dependent on the verb examining noun phrase constituents for syntactic and conceptual agreement with the pronoun. When a noun phrase

constituent is found that agrees with the pronoun, DEPSEARCH replaces the pronoun in the constituent in which it occurs by that noun phrase constituent.

LOOKLEX is called with a noun phrase constituent which has a possessive pronoun modifying it and the verbal constituent on which the noun phrase constituent is dependent as its arguments. It sees if the head noun of the noun phrase constituent has been used in a closely preceding portion of the discourse by examining the TOK-1 relational link on the definitional list of the lexical entry for the head noun. If it has occurred in a constituent for a closely preceding portion of the discourse, this constituent is examined to see if it has possessive modification indicated by POSS or OF relations. If the constituent has a possessive modifier this modifier is examined for syntactic agreement with the possessive pronominal modifier of the argument noun phrase constituent. If the possessive modifier agrees with the possessive pronoun syntactically, LOOKLEX returns the modifier as its value. If no possessive modifier is found that agrees syntactically with the possessive pronoun, then LOOKLEX returns NIL.

LOOKLEX1 is called with a noun phrase constituent and the verbal constituent on which it is dependent. It checks the lexical entry of the head noun of the noun phrase constituent for a TOK-1 link to another noun phrase constituent. If this noun phrase constituent is related to a verb which is within a prescribed number of constituents from the argument

verbal constituent on the list of verbal constituents, the noun phrase constituent is returned as the value of LOOKLEX1. If no occurrence of the head noun is found within the prescribed distance, LOOKLEX1 returns NIL.

LOOKMOD is given a noun phrase constituent, a third person pronoun, and the constituent in which the pronoun occurs as its arguments. It examines all of the noun phrase constituents which are dependent on the argument noun phrase constituent for syntactic and conceptual agreement with the pronoun. If a noun phrase constituent is found that agrees with the pronoun, LOOKMOD returns it as its value. Otherwise LOOKMOD returns NIL.

LOOKSYN is called with a noun phrase constituent as its argument. It searches the main topics list for a constituent whose head noun is synonymous with the head noun of the argument noun phrase constituent by examining the lexical entries of the head nouns of constituents on the main topics list for one with an EQUIV relation with the head noun of the argument noun phrase constituent. If such a constituent is found on the main topics list, LOOKSYN returns it as its value. Otherwise, LOOKSYN returns NIL.

LOOKSUP is called with a noun phrase constituent as its argument. It searches the main topics list for a constituent whose head noun is in a SUP relation with the head noun of the argument noun phrase constituent. If such a constituent is found on the main topics list, LOOKSUP returns it as its

value. Otherwise, LOOKSUP returns NIL.

LOOKTOPICS is called with a noun phrase constituent as its argument. It searches the main topics list for a constituent whose head noun is either lexically equivalent with the head noun of the argument noun phrase constituent or in some relation with the head noun of the argument noun phrase. If such an item is found, LOOKTOPICS adds the appropriate relational link pointing to the item to the definitional list of the argument noun phrase constituent, and the reverse relational link pointing to the argument noun phrase constituent to the definitional list of the constituent from the main topics list.

LOOKUP is called with a third person pronoun and the constituent in which it occurs as its arguments. It searches the current list of main topics for a noun phrase constituent which agrees in gender and number with the pronoun and fits into the conceptual context of the pronoun. If such an item is found, LOOKUP returns it as its value. Otherwise, it returns NIL.

LOOKVERB is called with a noun phrase constituent and the verbal constituent on which it is dependent as its arguments It examines the lexical entry of the head noun of the noun phrase constituent to see if it relates to a verbal form. If it does, LOOKVERB gets the verbal constituent which precedes the argument verbal constituent from the list of verbal constituents and sees if its verb is equivalent

as its arguments. ADDTO adds the antecedent to the relative clause constituent in the same relation to the verb of the relative clause constituent as the relative pronoun. It adds the inverse relational link to the antecedent constituent, and deletes the relations RELCL-1 and RELCL from the relative clause constituent and the constituent on which it is dependent. Consider the following call to ADDTO:

```
ADDTO(RELPRO RELCLCONSTIT ANTECEDENT)
RELPRO = who2
RELCLCONSTIT = G3 = (TOK sick2)(SUBJ who2)(FOR G4)(RELCL-1 G2)
ANTECEDENT = G1 = (TOK man2)(DET a1)(SUBJ-1 G2)
G2 = (TOK die2)(SUBJ G1)(RELCL G3)
AFTER APPLICATION OF ADDTO:
G3 = (TOK sick2)(SUBJ who2 G1)(FOR G4)
G1 = (TOK man1)(DET a1)(SUBJ-1 G2 G3)
G2 = (TOK die2)(SUBJ G1)
```

REPLACE is called with a third person pronoun, the constituent in which the pronoun occurs, and the antecedent of the pronoun as its arguments. It replaces the pronoun in the constituent in which it occurs by its antecedent, and adds the inverse relation to the antecedent constituent. Consider the following call to REPLACE:

```
REPLACE(PRONOUN CONSTITUENT ANTECEDENT)
PRONOUN = it1
```

CONSTITUENT = G1 = (TOK grow1)(MOD larger1)(SUBJ it1)(IF-1 G2)

ANTECEDENT = G3 = (TOK cell1)(DET the1)(SUBJ-1 G2)

After application of REPLACE:

G1 = (TOK grow1)(MOD larger1)(SUBJ G3)(IF-1 G2)

G3 = (TOK cell1)(DET the1)(SUBJ-1 G2 G1)

REPLACE1 is called with an anaphoric constituent formed for a demonstrative adjective modifying a noun phrase and the antecedent of this anaphoric constituent as its arguments. REPLACE1 deletes the anaphoric constituent and replaces all relational links to it in other constituents by relational links to the antecedent constituent. It also adds the inverse relational links to the antecedent constituent. Consider the following call to REPLACE1:

REPLACE1(ANAPHCONSTIT ANTECEDENT)

ANAPHCONSTIT = G4 = (TOK book2)(DEMADJ this1)(SUBJ-1 G3)

ANTECEDENT = G2 = (TOK book2)(DET a1)(MOD small2)(OBJ-1 G1)

G3 = (TOK have1)(SUBJ G4)(OBJ G5)

AFTER APPLICATION OF REPLACE1:

G4 deleted

G2 = (TOK book2)(DET a1)(MOD small2)(OBJ-1 G1)(SUBJ-1 G3)

G3 = (TOK have1)(SUBJ G2)(OBJ G5)

BIBLIOGRAPHY

Celce, Marianne, and Olney, John C. "The Three General Functions of 'the' with Special Attention Given to Anaphora." Unpublished paper. Santa Monica, California: System Development Corporation, 1969.

Celce, M., and Schwarca, Robert M. "A Semantically Motivated Description of the Articles in English." Unpublished paper. Santa Monica: System Development Corporation, 1969.

Fillmore, Charles J. "The Case of Case." Universals in Linguistic Theory. Bach, E, and Harms, R.T., eds. New York: Holt, Rinehart, & Winston, 1968, pp. 1-91.

Jones, Karen Sparck. "Notes on Semantic Discourse Structure." Santa Monica: System Development Corporation Document, SP-2714, March 3, 1967.

Karttunen, Lauri. "Coreference and Discourse." Unpublished paper. Austin, Texas: University of Texas, 1968.

Karttunen, Lauri. "What Makes Definite Noun Phrases Definite?" The RAND Corporation, P-3871, June, 1968.

Lakoff, George. "Pronouns and Reference." Unpublished paper. Cambridge, Massachusetts: Harvard, 1968. (Mimeographed by Indiana University Linguistics Club.)

McCarthy, John, and others. Lisp 1.5 Programmer's Manual. Cambridge, Massachusetts: The M.I.T. Press, 1968.

Olney, John C., and Londe, Dave L. "A Research Plan for Investigating English Discourse Structure, with Particular Attention to Anaphoric Relationships." Santa Monica: System Development Corporation Document, TM(L)-3256, November 22, 1966.

Olney, J. C., and Londe, D. L. "An Analysis of English Discourse Structure, with Particular Attention to Anaphoric Relationships." Santa Monica: System Development Corporation Document, SP-2769, November 10, 1967.

Olney, J. C. "An Investigation of English Discourse Structure, with Particular Attention to Anaphoric Relationships." Santa Monica: System Development Corporation Document, TM-4229, March 5, 1969.

Ross, John Robert. "On the Cyclic Nature of English Pronominalization." To Honor Roman Jakobson. Vol. III. pp. 1669-1682. The Hague: Mouton & Co., 1967.

Simmons, Robert F., and others. "A Computational Model of Verbal Understanding." Santa Monica: System Development Corporation Document, SP-3132, April 30, 1968.

VITA

Sharon Baranofsky was born in Houston, Texas, on August 25, 1947, the daughter of Bernice Commander Baranofsky and Stanley Baranofsky. After completing her work at Lamar High School, Houston, Texas, in January, 1965, she entered the University of Houston, in Houston, Texas. She transferred to the University of Texas at Austin in September, 1966. She received the degree of Bachelor of Arts, majoring in Mathematics and minoring in Physics, in May, 1968. In June, 1968, she entered the Graduate School of the University of Texas at Austin. During the summer and fall of 1969 she was employed as a programmer for a research project under the direction of Dr. R. F. Simmons in the Department of Computer Sciences at the University of Texas at Austin.

Permanent Address: 3408B Cedar
Austin, Texas 78705

This thesis was typed by Pamela R. Kuhn.