CANADIAN THESES ON MICROFICHE

THÈSES CANADIENNES SUR MICROFICHE

National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

Canada

NL 339 (r. 86/09)

THE UNIVERSITY OF ALBERTA

COGNITIVE ASPECTS OF ANAPHORA JUDGMENT AND RESOLUTION

by

(C) RONALD H. SMYTH

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

DOCTOR OF PHILOSOPHY

IN

PSYCHOLINGUISTICS

DEPARTMENT OF LINGUISTICS

EDMONTON, ALBERTA

SPRING, 1985

For Chloe, St. Debbie

ABSTRACT

English pronominal anaphora is examined from three perspectives: (i) the role of processing strategies in grammaticality judgment; (ii) models of on-line resolution; and (iii) post-resolution effects on priming and activation.

On the basis of an overview of formal approaches to anaphora, a detailed analysis is presented of non-syntactic factors influencing coreference judgments by nonlinguists. It is argued that while naive speakers apply comprehension strategies during the judgment task, linguists are more likely to apply production frequency criteria. In a judgment study involving 33 examples of blocked coreference drawn from the literature, speakers allowed coreference on 55% of the items. Discussion focuses on methodological aspects of data collection for anaphora studies.

The experimental literature on anaphora resolution is also reviewed in detail. The exhaustive serial search model is challenged, and a new model incorporating on-line exploitation of syntactic and semantic cues is offered. This model implies that the "gender effect" (Springston, 1975; Kaplan, 1981) is an artifact of the use of verification tasks to study the resolution process, and an experiment separating reading times from verification times shows that the effect is attributable entirely to the latter task. A model of anaphora verification further predicts that the effect should appear only in

free anaphora, and is supported.

Finally, post-resolution activation effects are studied in a probe recognition task. It is shown that differences in activation for an antecedent and a word in the same proposition are due to selective partial deactivation of the "companion" word. This is interpreted as evidence for propositional representation and proposition-based retrieval even when surface codes are still available shortly after sentence processing. It is argued that ordinary reading tasks are appropriate for probe reaction studies, even with subject-controlled single sentence displays followed by a subject-controlled probe onset.

ACKNOWLEDGEMENTS

This thesis is an attempt to answer the question, "How do pronouns work?". The fact that I felt it both necessary and possible to answer it in two ways shows that my mentors at the University of Alberta have provided the kind of guidance I was seeking when I first learned of their commitment to the reunification of linguistics and psychology. The fact that the linguistic arguments are phrased in terms of cognitive processes, and that the psychological arguments stem in part from a linguistic analysis, is evidence of their influence.

I am especially indebted to Professors Bruce Derwing and Gary Prideaux, who taught me to explore the psychological implications of linguistic theories. They showed me that language is both simpler and more complicated than other linguists tend to think.

Of my many friends in Edmonton and Calgary, Tracey Derwing stands out as having made the greatest personal contribution to the completion of this work. The importance of her constant support, encouragement and prodding is inestimable. Her dog, however, I can live without.

Two other marvelous duos, Peter Assmann and Albi Calman, and Wendy Rollins and Graham Peacock, also extended friendship, advice, meals, and, in the latter case, even a home for the summer.

I would also like to thank Dr. John Darbyshire and Mrs. Jenny White of Hotel Dieu Hospital, Kingston, for allowing me to use hospital

facilities and staff in the running of my experiments. Also at Hotel Dieu, Luis Zegarra is the man most responsible for the completion of Experiments 2 and 3. Without the countless hours he spent trying to meet my changing programming needs on unfamiliar equipment, I could not have undertaken this type of research.

Sylvia Hains of the Department of Psychology at Queen's University put in long hours over this past summer running my statistical analyses and at the same time guiding me through yet another new operating system. Her patience and sense of humour were greatly appreciated. I am also grateful to that Department for providing the computing facilities and office space, and to my students for commenting on earlier versions of my ideas.

Finally, I would like to thank my other committee members for their valuable comments: Dr. Matthew Dryer of the Department of Linguistics, Dr. Peter Dixon of the Department of Psychology, and Professor Jim McCawley of the University of Chicago.

TABLE OF CONTENTS

TABLE OF CONTENTS continued

TABLE OF CONTENTS continued

TABLE OF CONTENTS continued

TABLE OF CONTENTS continued

LIST OF TABLES

List of Figures

CHAPTER ONE

OVERVIEW

1.1 Introduction

The problem of how English speakers are able to compute pronoun-antecedent relations, part of the broader anaphora resolution question, is receiving a good deal of attention from researchers in natural language syntax, discourse analysis, cognitive psychology, and artificial intelligence alike. Such an expenditure of time and energy might seem extravagant to the layman, for whom the process is entirely automatic, but attempts to model anaphora resolution have shown that it involves a surprisingly large number of morphological, lexical, syntactic, semantic, pragmatic, discourse, and suprasegmental variables. Moreover, identifying the many linguistic influences on coreference is only half the battle; one must also account for the implementation of these cues during the antecedent search process, in a manner which is compatible with existing models of word recognition, parsing, semantic integration, and memory. Anaphora resolution therefore provides a formidable challenge for the development of workable human and machine models of language comprehension.

This study begins, in Chapter II, with an extensive review and

elaboration of formal linguistic treatments of pronominal anaphora, followed by a discussion, in Chapter III, of several processing strategies which appear to influence whether or not speakers will judge coreference to be possible for a particular sentence. These strategies suggest that coreference judgments will be highly contaminated by factors irrelevant to the goal of specifying a speaker's output constraints on anaphoric relations. This hypothesis is tested, and confirmed, in the judgment experiment reported in Chapter IV.

Chapter V reviews the information processing literature on anaphora resolution, examining the factors governing the many search, assignment, activation, and integration processes triggered when a pronoun is encountered in text. This includes such questions as the type of search involved (serial vs. parallel; exhaustive vs. self-terminating), when resolution begins (during processing of the pronoun itself, or later in the clause, sentence or discourse) and ends (again, during pronoun processing or later), the type of memory model required (semantic activation of antecedents), search domains (effects due to clause or sentence boundaries, foregrounding, and the number of potential preceding and following antecedents), search priorities (which potential antecedents are checked sooner or more quickly), search direction (forward vs. backward), and many other related issues. Chapters VI and VII present two further experiments concerning pronoun processing. Experiment 2 compares search processes for antecedents which both precede and follow the pronoun, and for both syntactically permitted and disallowed antecedents. The results of this study are interpreted from the perspective of competing top-down and bottom-up models of the comprehension process. Experiment 3

demonstrates that even in very short-term searches (through an immediately preceding clause), a pronoun activates both its antecedent and NP concepts in the same clause, a result which is not entirely accounted for by Kintsch's (1974) model of the propositional representation of text in memory.

The final chapter consists of a general discussion.

1.2 An example

Consideration of the following pair of sentences will assist the reader in gaining a preliminary grasp of the complexity of anaphora resolution.

1. Vietnam is currently seeking advice from other countries on agricultural matters.
2. Due to its enormous population, China has the kind of experience in food production that Vietnam needs.

How does a reader or listener determine that the intended antecedent of its is China? Resolution appears to depend on at least the following factors:

(i) its is singular and neuter (its ≠ countries): MORPHOLOGY

(ii) its may refer to an NP either within or outside the sentence in which it appears (its = either Vietnam or China): SYNTAX

(iii) Vietnam is the initial topic of discourse, and the only country mentioned when the pronoun is encountered (its = Vietnam is a preferred reading): DISCOURSE STRUCTURE.

(iv) neither experience nor food production may be said to have a population (its ≠ experience or food production): LEXICAL SEMANTICS

(v) the population of China is enormous, while the population of Vietnam is not (its ≠ Vietnam): WORLD KNOWLEDGE

These are but some of the cues to correct resolution to be discussed in the following chapters.

Such an analysis also raises a variety of interesting questions concerning processing. For example, if resolution is attempted on-line, (2) should be a garden-path sentence. Moreover, one would like to be able to specify at what point an initial hypothesis (its = Vietnam) begins to disintegrate: is processing disrupted immediately in response to incoming information concerning the population of the asserted antecedent, or are such complexities deferred to a later stage, such as the next clause boundary? Further questions concern the nature of the search: Are only NP's scanned? Is it the contents of working memory that are canvassed? Does distance from the pronoun have an effect? Are backward searches easier than forward searches? What is the time required to conduct the search?

This single example suggests that the study of anaphora resolution is not only a fruitful enterprise, but also one which gives us fascinating insights into the detailed structure of language comprehension.

CHAPTER TWO

FORMAL ACCOUNTS OF CONSTRAINTS ON ANAPHORA

2.1 Introduction

Anaphoric processes were given scant treatment by linguists prior to the 1960's. Early analyses (e.g., Bloomfield, 1933; Jespersen, 1909-1949; Hockett, 1955; Gleason, 1955) were concerned mainly with the morphological aspects of pronoun systems. With respect to the syntactic phenomena, they made only passing reference to the fact that pronouns "replace" or "stand for" full NP's (see Crymes, 1968). Examination of syntactic constraints was begun by Lees and Klima (1963), and certain of the resolution biases for otherwise free anaphora were recognized by Jackendoff (1973). However, as recently as 1973, Quirk, Svartvik and Greenbaum say of contextual factors in resolution only that "we interpret the appropriate substitutions from the content of the sentence" (p. 294).

Syntactic analyses attempt to characterize such restrictions in terms of one or two principles of interpretation from surface phrase structure; they differ mainly in their definitions of the

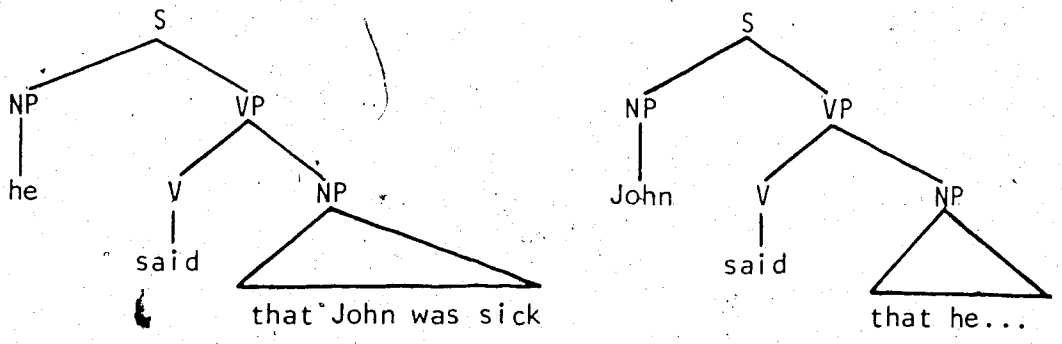
dominance relations on which these principles depend. Much of this work constitutes an attempt to clarify the role of surface phrase structure relations such as precedence and command in blocking coreference between personal pronouns and candidate antecedents (e.g., Chomsky, 1980, 1981; Langacker, 1966, 1969; Reinhart, 1976, 1981; Wasow, 1979). Wasow provides this simplified account:

> Given an NP and a definite pronoun in the same sentence, the NP may serve as the antecedent for the pronoun unless...the pronoun is to the left of the NP and is less deeply embedded than the NP (p. 61).

Coreference is thus blocked for he and John in (3) below, where he must to refer to some other, unspecified antecedent, and free in (4), where, given an appropriate context, he may refer either to John or to some other NP:

3. *He said that John was sick.

4. John said that he was sick.



(Tree structures are simplified; asterisks indicate blocked coreference between underlined NP's)

According to the most popular classification scheme, the blocked co-reference is said to run backward from John to he in (3), while (4) is termed free forward coreference.

Interpretive theories within the transformational paradigm date

from at least Jackendoff (1968), and in general have attempted to provide a comprehensive account of sentence-level syntactic constraints which block coreference between specified NP's. Because of the nature of the data on which the earlier accounts were based, this was for several years referred to as the Backward Anaphora Constraint. Later recognition of the existence of blocked forward anaphora has led to the demise of this term.

It should be emphasized that in syntactic accounts, lexical, semantic, pragmatic, discourse, suprasegmental, and processing variables are treated as nuisance variables modulating coreference interpretation within a set of fundamental syntactic constraints. This view is far from unanimous: more pessimistic accounts (e.g., Bily, 1977; Bolinger, 1977, 1979) view the entire syntactic approach as a red herring, and claim that counter-examples are so numerous as to cast serious doubt on the very assumption that anaphora is a syntactic process.

2.2 Definitional preliminaries

Perhaps a good deal of potential terminological confusion can be forestalled by examining here the various uses of the term anaphora by linguists.

Bloomfield (1933) introduces the terms anaphora and antecedent in a chapter entitled 'Substitution phenomena':

> ...the form for which substitution is made has occurred in recent speech. Thus, when we say Ask that policeman, and he will tell you, the substitute he means among other things that the singular male substantive expression which is replaced by he, has been recently uttered. A substitute which implies this, is an anaphoric or dependent substitute, and the recently-uttered replaced form is the antecedent. (p. 249)

Thus for Bloomfield, a pronoun must have an explicit NP antecedent, which precedes it in discourse (Bloomfield does not discuss backward anaphora), and which may or may not be in the same sentence as the pronoun.

Waterhouse (1963), on the other hand, defines anaphora as a purely inter-sentential phenomenon: an anaphoric pronoun is one "without identified referent in the same sentence" (p. 46), and until quite recently, transformationalists defined anaphora as a purely intra-sentential phenomenon. Dougherty (1969) states that a pronoun

> ...with anaphoric interpretation...can be understood as being coreferential with some other noun phrase in the sentence (p.)

In transformational terms, a sentence containing a free pronoun is often said to be ambiguous between the anaphoric and non-anaphoric readings (but cf. Lasnik, 1976), while a sentence with blocked coreference is said to have only a non-anaphoric reading. Thus, clauses containing reflexive and reciprocal pronouns normally have only an anphoric reading. This terminology, according to which anaphoricity is a property attributed to entire sentences, is problematic for cases such as

5. John told Fred to hit him.

in which the pronoun is ambiguous between the anaphoric and non-anaphoric readings with respect to John, but has only a non-anaphoric reading with respect to Fred. That is, anaphoricity is not a property of the sentence as a whole, but only of the syntactic relations between NP's.

Halliday and Hasan (1976) distinguish between endophoric and

exophoric reference, reserving the latter term for situational ("deictic" or "pragmatic") reference, and the former for reference within a text. Endophoric reference is further categorized according to whether the antecedent appears earlier (anaphoric) or later (cataphoric) in the text, but these terms do not distinguish between inter- and intra-sentential coreference.

Most recently, Chomsky (e.g., 1980, 1981) and others working within the Government and Binding framework have made a different distinction between anaphoric and pronominal expressions. The former include reflexives and reciprocals, for which coreference is syntactically determined, and the latter include the personal pronouns regardless of whether coreference is free or blocked in particular sentence in which they appear.

In the following pages, I shall take the term anaphora to mean a relation of either possible (in the case of simple pronouns) or necessary (for reflexives and reciprocals) coreference between two NP's, defined in terms of linguists' or native speakers' judgments and usage. Where convenient, I adopt Dougherty's distinction between anaphoric and non-anaphoric sentences, as well as Halliday and Hasan's finer distinctions.

A final point of definition concerns the nature of reference itself, whose explication is obviously beyond the scope of this work. Following Lyons (1977), I shall occasionally adopt what has become colloquial usage among linguists, according to which a pronoun is said to refer to its antecedent. Of course, this is not strictly correct, since the essential characteristic of (non-deictic) pronouns

is that they share a referent ("co-refer") with their antecedent. The antecedent itself, being a linguistic expression, is not the referent.

2.3 Introduction to syntactic accounts

As a first step toward constructing a model of anaphora resolution, we turn now to attempts at characterizing the linguistic environments in which coreference is possible, impossible, probable, or indeterminate. One source of such information is the literature dealing with syntactic constraints on pronominal coreference.

In examining this literature, two caveats are in order. First, none of the theories currently available is "correct": all make a certain number of wrong coreference predictions. This fact in itself bodes poorly for the popular notion that coreference options are governed by a single principle of interpretation from surface phrase structure. Secondly, there is much disagreement as to the data to be accounted for. While some of these disagreements are discussed openly in the literature, I will demonstrate that judgments of coreference options are exceedingly susceptible to a variety of influences, and further, that such indeterminacy is an unavoidable consequence of the fact that anaphora is at once a syntactic, semantic, and discourse phenomenon.

2.4 Reflexives

In this section I present several miscellaneous facts about coreference constraints on reflexives in order to determine the viability of the interpretive approach to describing them.

The simplest statement of this is the Clausemate Constraint, first proposed (as a transformational rule) by Lees and Klima (1963).

Lees and Klima noted that when two NP's in the same simplex sentence are coreferential, the rightmost NP is obligatorily a reflexive pronoun of appropriate gender, number and person:

6. Albi showered herself with gifts.

Conversely, a reflexive pronoun obligatorily has a clausemate as its antecedent. Thus, the ungrammaticality of (7):

7. *Peter showered herself with gifts.

is attributable not to a lack of antecedent for herself, but to the assertion of coreference between the gender-mismatched Peter and herself.

Lees and Klima claimed that the constraint holds without exception only when the reflexive functions as a direct or indirect object; in other positions the constraint is relaxed with certain verbs:

8. John ignored the oil on him.

9. John smeared the oil on himself.

However, this is an overly simplistic account, since the PP's of (8) and (9) differ in other ways. Since the PP of (8) can be viewed as a reduced relative clause, the use of a non-reflexive can be attributed to the fact that John and him are not in the same simplex sentence, as required by their statement of the rule.

Nonetheless, reflexives are, in most cases, in complementary distribution with simple pronouns, which cannot be coreferential with a clausemate:

10. *Peter saw him in a mirror.

11. *Patti bought her a new wading pool.

(Note that the asterisks indicate only that the coreferential readings are impossible, and not that the sentences are otherwise ungrammatical).

Ross (1970) notes a class of exceptions to the Clausemate Constraint: a reflexive pronoun may, in some cases, have its antecedent in the next highest clause if and only if there is no intervening clausemate NP which agrees with it in animacy. This is true of the so-called "picture noun" sentences such as

12. Tom believes that there is a picture of himself hanging in the post office.

In addition, coreference can skip over the picture noun only to the next highest NP which agrees in animacy:

13. *Tom believes that Sheila saw a picture of himself.

In a sense, then, a picture noun plus PP modifier behaves as though the head of the PP were a clausemate of the nearest higher non-picture NP.

Two additional facts are relevant to this account. First, consider the case of two potential antecedents in the higher clause, both matching the reflexive in animacy:

14. The waitress showed Sheila a picture of herself.

Jackendoff (1972) offers similar examples of the apparent tendency to interpret such sentences with resolution to the subject. That this is merely a bias is demonstrated by sentences such as

15. The waiter showed Sheila a picture of herself.

in which a gender mismatch with the subject allows object resolution. In addition, it is interesting to ask how a speaker would express (14)

with herself = Sheila: would one simply use the ambiguous form, hoping to clear up the vague reference later, or would one use the simple pronoun?

16. The waitress showed Sheila a picture of her.

Secondly, consider the case of an extra intervening inanimate noun: can coreference skip both?

17. Lorne noticed that the slide showed a sketch of him/himself.

I have encountered mixed reactions to such sentences. Most speakers say that both are possible, but they disagree as to which is better. Those who have a fairly strong preference for the simple pronoun (myself included) find the following cases perfectly acceptable:

18. Lorne told Jack that the slide showed a sketch of himself.

19. Lorne told Jack that the slide showed a sketch of him.

This minimally contrasting pair suggests that despite the split opinions over (17), coreference can indeed skip over two inanimate NP's, and that the reflexive/simple pronoun contrast is functional when only a resolution strategy can accomplish interpretation: reflexives bias the interpretation to the higher subject; non-reflexives bias interpretation to the non-subject. I do not believe that this has been discussed in the literature on reflexives.

The fact that speakers disagree on this point is in itself rather interesting, for it suggests that the rarity of picture noun sentences, and especially of these specially chosen examples, can result in non-application of the interpretive principles governing the simpler cases. The existence of speakers who reject the distinction shown in (18) - (19) raises the possibility that speakers are not necessarily able

to generalize interpretive principles to unusual cases. If this is so, the validity of collecting naive speakers' judgments on crucial cases is placed in jeopardy.

Chomsky (1981) shows that in certain subjectless complement clauses, the Clausemate Constraint is relaxed to allow coreference with an NP in a higher clause:

20. John appears to be enjoying himself.

His trace theory analysis, however, remains compatible with the constraint if one assumes that the reflexive pronoun is coreferential with the trace left by movement of John out of the lower clause:

21. John appears t to be enjoying himself.

Finally, Ross (1970) presents an example in which an animate NP may be the antecedent for a pronoun in a lower clause, even with intervening animate NP's:

22. Albert accused me of having tried to get his secretary to tell you that physicists like himself were hard to get along with.

Only if the intervening NP's match the reflexive pronoun on all relevant features does complete ambiguity result:

23. Albert accused Fred of having tried to get Mark to tell Michael that physicists like himself were hard to get along with.

Despite the possibility of coreference with any of the proper names, only Albert and Michael seem to be likely antecedents. This can be attributed to the pragmatics of the verb accuse: he = Fred is pragmatically bizarre, since criticizing oneself is not the sort of reprehensible act of which one might normally be "accused" (a similar argument holds for him = Mark). On the other hand, the reading is not

entirely excluded, since Fred could have some complex motive for wanting to convince someone else of his own character flaws - perhaps he wishes to get out of a prior agreement to work with Michael. It would appear that the amount of inferencing required in order to construct such a context somehow, precludes the two marked readings. Of course, this too brings forth a number of questions that are relevant to both human and machine comprehension models: how much inferencing is actually done before a reading is adopted? Would the non-preferred reading still be more difficult to arrive at in a context in which it is plausible? What type of memory model could account for such context effects?

To summarize, then, the Clausemate Constraint on the interpretation of direct and indirect object reflexive pronouns provides a syntactically based, deterministic resolution strategy: the pronoun is obligatorily coreferential with its clausemate, subject to certain lexical and syntactic biases.

Since I have approached reflexives from the standpoint of interpretive principles, I should now reiterate that Lees and Klima originally formulated the constraint as a transformation which replaced a lexical NP by a reflexive pronoun. This will be clarified in the following sections.

2.5 Simple pronouns

The third person pronouns (he, she, it, they (nominative); his, her, its, their (genitive); him, her, it, them (accusative)) are not subject to deterministic resolution principles. For example,

although a simple pronoun cannot have a clausemate antecedent:

24. *John saw him in the mirror.

the antecedent is unspecified. Perceiving that coreference is blocked is only one step toward resolution; other structural principles governing the free/blocked distinction are the topic of the following sections.

2.6 Pronominalization and reflexivization as transformations

Lees and Klima (1963) proposed that pronouns are introduced by a transformational rule of Pronominalization (or Reflexivization) which replaces an underlying full NP with a pronoun under morphological identity. Chomsky (1965) found it necessary to alter this approach in order to conform with the Katz-Postal Hypothesis (Katz and Postal, 1964), which states that meaning is interpreted from deep structure; this accounts for the fact that pairs such as

25. John hurt John.

26. John hurt himself.

may differ in interpretation: (25) is ambiguous between the anaphoric and the non-anaphoric reading, but (26) has only an anaphoric reading. Since in the Lees-Klima formulation the deep structure of (26) is roughly (25), the theory would incorrectly predict that (26) is also ambiguous. Chomsky therefore proposed that deep structures be altered so as to allow the generation of referential indices along with underlying NP's. The two readings of (25) would then derive from distinct deep structures differing in whether or not the two occurrences of John are coindexed. By adding the condition that

Pronominalization and Reflexivization include a referential identity condition, one can derive (26) unambiguously from the deep structure in which the NP's are coindexed.

Over the next few years, however, it became apparent that the transformational analysis was untenable for other reasons. Some of the main arguments for the alternative interpretivist theory (in which coreference is determined from derived phrase marker configurations) will be presented here.

Perhaps the best-known (but no longer accepted; see Perlmutter and Soames, 1979) argument against the transformationalist position is the so-called Bach-Peters paradox (Bach, 1970), which shows that if the identity condition is to be maintained for co-indexing, the assumption that phrase markers must have a finite number of branches is violated. Consider (27):

27. The pilot who shot at it hit the mig that chased him.

Here, it would be coindexed with the mig that chased him, and him would be co-indexed with the pilot who shot at it. However, each of these full NP's contains a pronoun which must be co-indexed with another NP which contains a pronoun, and so on, ad infinitum. Under the assumption that infinite deep structures are disallowed, the transformational analysis must be rejected.

A second argument (Lakoff, 1968, 1969) concerns pronominal epithets - NP's which (when unstressed) may be coreferential with a lexical antecedent:

28. I told Ms. Barham about the sink, but the stupid bitch wouldn't do anything about it.

Since these NP's seem to obey all the distributional restrictions

that apply to pronouns, one would expect a theory of anaphora to handle them with the same formal mechanism. However, the transformational treatment allows coreference only under deep structure identity; the Pronominalization rule would have to be altered so as to allow for substitution of epithets rather than pronouns in such cases. This poses a problem in that epithets also carry meaning, which cannot be interpreted from the later stage in the derivation. In other words, a transformational approach to the interpretation of pronominal epithets violates the Katz-Postal hypothesis.

A third set of arguments (Bresnan, 1970) is based on the interaction of Pronominalization with movement rules such as Particle Movement and Dative Movement. Since these rules apply cyclically, it is possible for an NP to be moved within a lower clause before an antecedent in a higher clause is "seen" by the Pronominalization rule, and for that NP to be pronominalized after it has been moved into a position where pronouns are disallowed:

29. The girl claims that John turned the girl away (Deep structure).

30. The girl claims that John turned away the girl. (Particle Movement)

31. *The girl claims that John turned away her. (Pronominalization).

The only way to prevent this would be to impose an ad hoc constraint preventing Pronominalization from applying to structures to which certain other transformations have already applied.

Finally, Dougherty (1969) argues that in treating anaphoric pronouns differently from non-anaphoric pronouns, the transformational analysis fails to account for the fact that for every sentence

with an anaphoric reading, there is also a non-anaphoric reading (Wasow, 1979, notes several counter-examples). Thus, the two readings of

32. John said he was sick.

have very different derivational histories (one with an underlying non-coreferential pronoun, the other with a co-indexed John), and the fact that (32) is ambiguous appears to be accidental. On the other hand, a theory which interprets coreference from surface phrase structure and generates all pronouns in deep structure would simply label (32) as a case of free anaphora and correctly interpret it as ambiguous between the anaphoric and the non-anaphoric readings. Since Chomsky (1971) had also argued that interpretation of meaning from surface structure is necessary in the case of topic and focus, this innovation was ultimately adopted, and the Katz-Postal Hypothesis laid to rest.

Further arguments against the transformational treatment may be found in Wasow (1979), Jackendoff (1968), and elsewhere.

2.7 Interpretive theories of anaphora

None of the above arguments against the transformational model poses a problem for an interpretive theory of anaphora. In interpretive theories, all pronouns are introduced by lexical insertion and coreference is determined on the basis of derived structural configurations. This solves the problems inherent in the transformational analyses: the Bach-Peters sentences are generated with pronouns in deep structure, and the problem of infinite recursion does not arise; the meanings of pronominal epithets are available in deep structure, and coreference interpreted at a later stage; and independent con-

straints prevent the application of the offending movement rules to block sentences such as (31).

As has already been suggested, a finding that all cases of free anaphora share a structural property not found in blocked cases would be of great interest, particularly if that property could be shown to be reflected in aspects of linguistic performance, such as processing complexity, language acquisition, or judgment data. The next three sections examine some specific proposals concerning such interpretive principles.

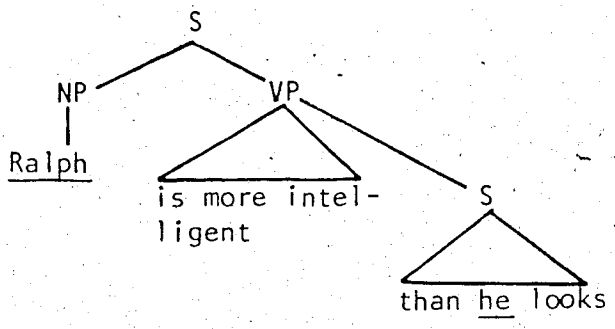
2.8 Precede-and-command

Langacker (1966, 1969) represents an early attempt to formalize the constraints on sentence-internal coreference in terms of the horizontal and vertical positions of pronouns and NP's in phrase structure. His constraint is that:

> A pronoun may not both precede and command its antecedent. ...A node A "commands" another node B if (1) neither A nor B dominates the other; and (2) the S-node that most immediately dominates A also dominates B. (p. 167)

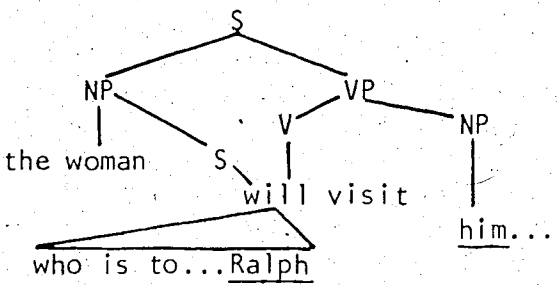
Langacker considered precedence and command to be the "primacy" relations relevant to anaphora, in that a pronoun which bears neither relation to an NP may be coreferential to (33), as may one which bears one of the relations (34,35). However, if both relations hold, coreference is blocked:

33. Ralph is more intelligent than he looks.



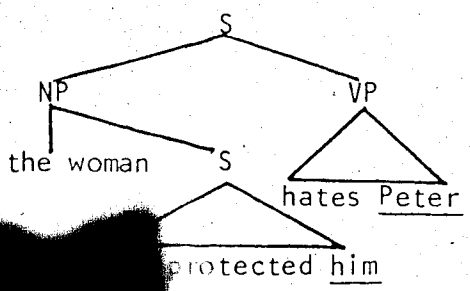
He neither precedes nor commands Ralph.

34. The woman who is to marry Ralph will visit him tomorrow.



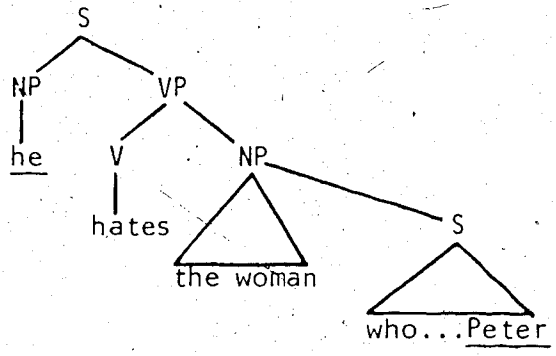
Him commands, but does not precede Ralph.

35. The woman who protected him hates Peter.



Hi[illegible]des, but does not command Pe[illegible]

36. *He hates the woman who rejected Peter.



He both precedes and commands Peter.

[illegible]hile the constraint accounts for a broad range of data, a num[illegible]r of counter-examples demonstrate that it is not entirely correct. First, it cannot account for all instances of acceptable backward core[illegible]ere[illegible]e:

37. Near him, Dan saw a snake. (Lakoff, 1968)

38. The chairman hit him on the head before the lecturer had a chance to say anything.

39. Irene sent letters to him while Fred was out of town.

40. I stayed with his wife while Fred was out of town.

In each case, the pronoun precedes and commands the antecedent, but the coreferential readings are possible.

Although Langacker failed to examine preposed adverbials such as (37), he implicitly accepts a prediction of ungrammaticality for (38) - (40) by starring (41):

41. * I knew him when Harvey was a little boy.

This judgment of noncoreference has been disputed by Reinhart (1981), and Carden (1978) cites several occurrences of such structures in spontaneous text. Thus the theory fails in these cases as well.

Precede-and-command also fails to block forward coreference of any kind:

42. *Near Dan, he saw a snake.

43. *How obnoxious to Ben's friends he is.

where the pronoun commands, but does not precede its antecedent.

It should also be noted that while the constraint blocks backward anaphora in some sentences which are otherwise handled by the Clausemate Constraint:

44. *He shot John.

it cannot account for forward blocking in similar structures:

45. *John shot him.

In (45) him does not precede John, yet coreference is blocked; the Clausemate Constraint is still necessary. Given the inadequacy of Precede-and-Command, one is left to speculate as to whether a single rule will emerge to replace it, or whether constraints on anaphora are

best handled by a larger number of rules.

2.9 Lasnik's Kommand analysis

Lasnik (1976) attempted to improve upon Langacker's analysis by providing an account of the acceptability of coreference in the case of genitive NP's which precede and command their antecedent (e.g., 40 above). This argument is based on Chomsky's (1969) observation that NP's can be cyclic nodes, in that they may head a domain for the application of transformations such as Passive:

46. The city was destroyed by the enemy.

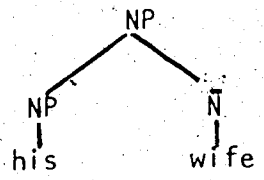
47. The city's destruction by the enemy...

48. The enemy destroyed the city.

49. The enemy's destruction of the city...

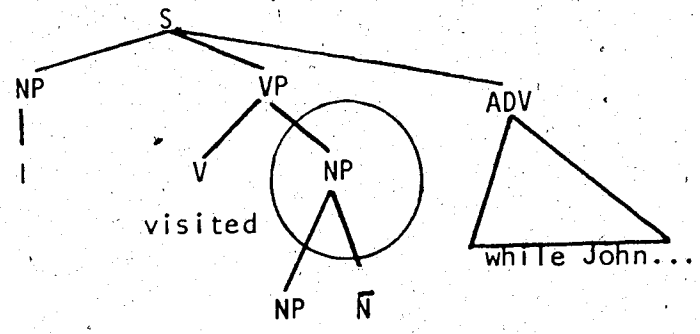
Lasnik observed that the dominance of a genitive NP by a higher NP node in structures such as (50):

50.



can be viewed as a barrier to command via the higher S node, as in (51):

51.



The circled NP node, being a cyclic node, dominates his but not

John, and no longer commands John under Lasnik's definition of the kommand relation:

52. A kommands B if the minimal cyclic node dominating A also dominates B. (p. 15)

Langacker's constraint is then altered as follows:

53. If NP_1 both precedes and Kommands NP_2, and NP_2 is not a pronoun, then coreference between NP_1 and NP_2 is blocked.

It is interesting to note that Lees and Klima (1963) saw a similar connection between cyclicity and pronominalization in their analysis of sentences such as (54):

54. Mary's father supported her / *herself.

which appear to be counter-examples to the Clausemate Constraint. They suggested that since "genitives in English are transformational in origin" (p. 22), the source of (54) is roughly (55):

55. The father$_{\#}$the father is Mary's$_{\#}$supported Mary.

with only one occurrence of Mary in the matrix sentence, and hence no reflexivization. Thus Lees and Klima realized that genitive NP's do not participate in coreference relations in the same way as other NP's in a main clause.

O'Grady (1983) deals with these exceptions in a partly semantic manner, claiming that genitives, as "peripheral" nominals with respect to the verb (subjects, objects and indirect objects are "core" nominals in his system), are permitted to be coreferential with the later NP because they do not violate his Precedence Constraint:

56. A core pronoun may not precede a peripheral antecedent.

It would appear that what is handled by dominance principles in the other theories is subsumed under the semantically-based distinction

between core and periphery.

We therefore have good evidence that genitives form a coherent class of exceptions to Langacker's constraint. However, Lasnik's analysis leaves a curious asymmetry in accounting for the grammaticality of (51), but not of similar structures such as (57):

57. I visited him while John was in Edmonton.

As we shall soon see, the grammaticality of structures such as (57) has been hotly debated. However, if (57) is grammatical even for only a subset of speakers, the cyclic nodes analysis cannot account for its existence. But this leads to two somewhat unsatisfactory conclusions: either the cyclic nodes analysis needs to be supplemented by yet another ad hoc specification of certain NP's which are not cyclic nodes, but which allow coreference in analogous structures, or the entire notion of cyclicity is not relevant to anaphora, and needs to be abandoned in favour of a set of exception rules to supplement precede-and-command. A third alternative, that command is simply the wrong relation, seems to be supported by the fact that the structure of (40) and of (57) share identical dominance relations despite the lack of cyclic nodes in the latter.

To carry the discussion one step further, note that the special NP-over-NP structure of Lasnik's cyclic nodes examples is also found in conjuncts; perhaps it is this abstract phrase marker configuration which affects the interpretation of coreference. For example, the following pair differ only in the presence or absence of the conjunct, yet (58) allows coreference, while (59) does not:

58. I told her and the producers that Meryl would be playing Sophie.

59. *I told her that Meryl would be playing Sophie.

(The argument becomes tenuous when we take into account the difficulty of collecting confident judgments on (58); see Chapter 4). On the other hand, conjuncts and genitives differ with respect to their interaction with the Clausemate Constraint: genitives, as we have already seen, are immune to that constraint, while conjuncts are not:

60. *John saw him in the mirror.

61. John saw his wife in the mirror.

62. *John saw him and Fred in the mirror.

Finally, even this argument is weakened somewhat by the fact that sentences such as (62) can be found in nonstandard dialects, and, I believe, in the language of children. This suggests that there may indeed be a link between conjuncts and genitives with respect to anaphora. The question of whether (40) and (57) allow coreference for the same reasons is thus unclear, but if that is indeed the case, the Kommand analysis cannot capture the generalization.

2.10 Reinhart and c-command

Reinhart (1976, 1981), while rejecting the precede-and-command analysis, nevertheless maintains that anaphora restrictions can be accounted for by a single surface interpretive principle. In addition to handling the blocked forward and free backward exceptions to precede-and-command, her constraint is also intended to account for certain asymmetries between coreference options for subjects and non-subjects, as in (63) and (64):

63. * In front of Dan, he saw a snake.

64. In front of Dan, I saw his snake.

This asymmetry poses a problem for any analysis incorporating command or Kommand, since everything c(k)ommanded by a subject is also c(k)ommanded by the object. However, simply restricting the application of some version of the Langacker analysis would be insufficient, since non-subjects can also be excluded from coreference, at least according to Reinhart's judgment on (65):

65. *I'm willing to give him two grand for Ben's car.

One very serious problem to be approached in these chapters is that of disagreement among linguists as well as among naive speakers as to which sentences in fact disallow coreference. Notice that Reinhart's analysis is largely justified by judgments such as (65); I will soon demonstrate that this and many other of her judgments are suspect, at least in terms of speaker's ability to judge the grammaticality of coreference in the crucial examples.

Reinhart's proposal is that

66. A given NP cannot be interpreted as coreferential with a distinct nonpronoun in its c-command domain. (p. 617)

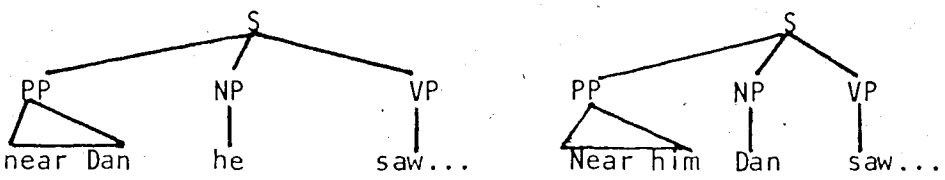
where c-command is defined as follows:

67. Node A c(onstituent)-commands node B iff the branching node α_1 most immediately dominating A either dominates B or is immediately dominated by a node α_2 which dominates B, and α_2 is of the same category type as α_1. (p.612)

C-command is simply the converse of Klima's (1964) in construction with relation, which Culicover (1976) used to account for free backward pronominalization (backward pronominalization is permissible if the NP is not in construction with the pronoun). Reinhart's insight is

that c-command is equally applicable to forward pronominalization, as in (68) and (69):

68. * Near Dan, he saw a snake. 69. Near him, Dan saw a snake.



In (68), Dan does not c-command he, but he c-commands Dan; in (69) the opposite is true. In (68), Dan is in the domain of he (where domain of a node refers to all the nodes that it c-commands); coreference is blocked. In (69), however, he is in the domain of Dan, but is not a pronoun; coreference is allowed. Notice in particular that the c-command formulation makes no reference to precedence.

As further examples, consider again the tree structures of (35) and (36) (page 21). In (35), no c-command relation exists between the pronoun and its antecedent, and restriction (66) does not block coreference. In (36), he c-commands Peter, and coreference is blocked by (66).

Since c-command can be shown to subsume the command relation, the only points of divergence between Reinhart's analysis and those of Langacker or Lasnik involve sentences in which (a) A c-commands B and B precedes A (since forward coreference is blocked, an impossibility under the other analyses), and (b) A precedes and commands B, but does not c-command it (such that only c-command correctly allows coreference). The former has already been illustrated in (68), where A = he and B = Dan (such sentences with preposed PP's are discussed

in detail below); the latter involves sentences such as (70), which Langacker judged to be ungrammatical:

70. We fired him, since McIntosh's weird habits had finally reached an intolerable stage.

Notice that (70) has the same structure as (57), which although superficially similar to the genitive sentences discussed by Lasnik, could not be accounted for by his theory; this formed the basis for the argument that cyclic nodes have no bearing on the issue. The fact that (70) and related structures are all predicted to allow coreference is encouraging for c-command, despite the matter of Langacker's disagreement with the judgment. One potentially serious objection to the constraint is the sheer number of questionable judgments to be found in Reinhart's work. Readers who have only a passing familiarity with c-command are invited to decide which of the following of Reinhart's sentences (some of them adapted) are free, and which are blocked:

71. For Irene's husband, she would give her life.

72. Fond of Donna's car though she is, I like it even more.

73. In Ben's film of Rosa, she found a scratch.

74. For her new glasses, how much did she pay?

75. In John's apartment, he smokes pot.

76. In Carter's home town he spends his most pleasant hours.

77. Michael's problems, he won't talk about.

78. When Lorraine finishes school, she promised Jim to go to Paris.

Some readers will be surprised to learn that all of these are starred by Reinhart. In a brief survey of naive speakers, I found such massive disagreement, not only with Reinhart but among subjects, that

either the theory or the tradition of collecting speakers' judgments must be thrown into question. This matter is discussed in deeper detail in Chapter IV.

2.11 Chomsky's binding conditions

Chomsky (1980, 1981) has adopted the c-command restriction in formulating his three Binding Conditions, which can be paraphrased as follows:

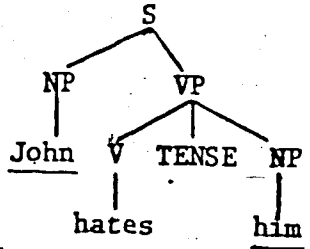
(i) A reflexive or recriprocal must be coreferential with any NP c-commanding it within the minimal S or NP that contains a V, A, N, P, or TENSE node which c-commands the reflexive or reciprocal.

(ii) A personal pronoun may not be coreferential with any NP meeting the conditions outlined in (i).

(iii) Lexical NP's may not be coreferential with any other c-commanding NP.

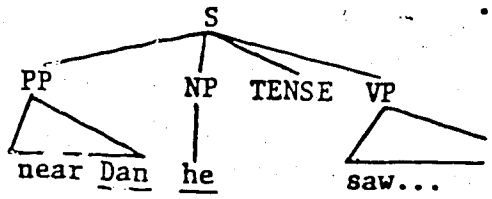
Conditions (ii) and (iii) are most relevant for our purposes. Consider the effect of these conditions in the following familiar examples:

79. *John hates him.



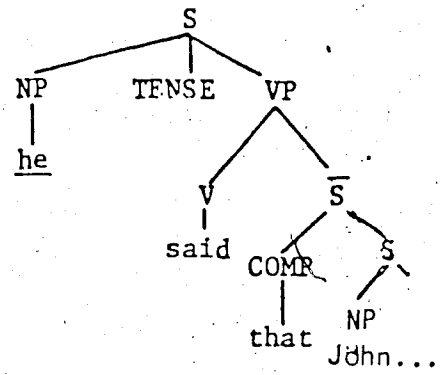
where John c-commands him within S which contains TENSE, which c-commands him.

80. *Near Dan, he saw a snake.



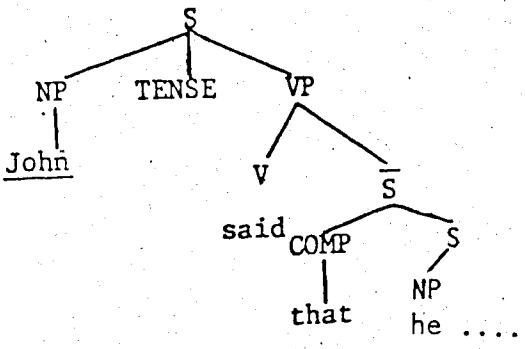
where he c-commands Dan within S, which contains P, which c-commands Dan.

81. *He said that John was sick.



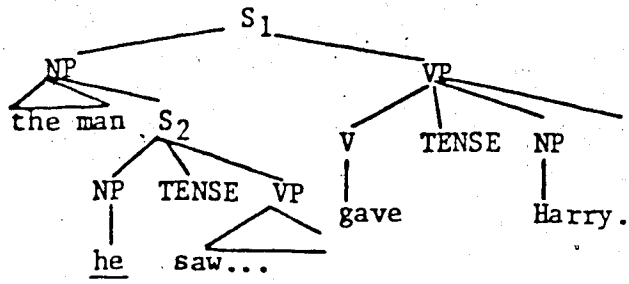
where no NP c-commands he within S; John may not be coreferential with he, which c-commands it.

82. John said that he was sick.



no NP c-commands John; condition (iii) does not apply. The minimal S containing TENSE, wich c-commands he, is S; no NP c-commands he within S.

83. The man he saw downtown gave Harry a gift.



no constraints apply: neither Harry nor he c-commands the other by the conditions in (i).

Thus the binding conditions cover the same range of data as Reinhart's original analysis. One significant difference is that condition (ii) accounts for the Clausemate Constraint as well as other cases of blocked anaphora; in Reinhart's model, the Clausemate Constraint was not accounted for. On the other hand, both conditions (ii) and (iii) are

necessary to account for the full range of blocked and free cases - there is no single interpretive principle.

In subsequent discussion, it is therefore assumed that difficulties in interpreting the c-command constraint in performance terms apply equally to the interpretation of binding conditions.

2.12 C-command, adverbial scope, and process interpretations

One of the most notorious problems with c-command involves its handling of preposed adverbials. In this section I shall attempt to shed some light on this issue by considering how c-command might be implemented as a coreference decision routine in a machine natural language parser. The purpose of this exercise is not to build a straw-man argument against c-command by demonstrating its failure to model human comprehension processes. On the contrary, in addition to helping clarify certain questions of syntax versus semantics, the machine implementation problem will provide a useful example of the very different goals of product and process models.

Reinhart notes sentence pairs purported to differ in coreference options because of differences in the VP which give rise to differences in adverbial scope. Consider these examples:

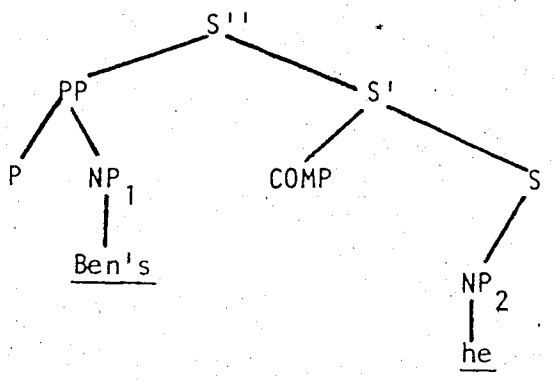
84. In Ben's office he is an absolute dictator.

85. * In Ben's office he placed his new brass bed.

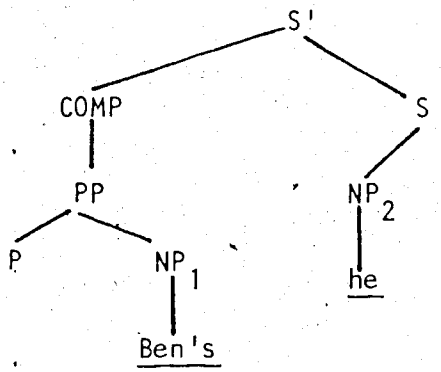
Although Wasow (1972) and Bickerton (1975) have suggested that the differences are attributable to performance constraints or presupposition respectively, Reinhart claims that the differences are due to purely structural factors: the PP of (84) is attached at S, while

that of (85) is attached at VP. In surface structure, this is represented by having the VP-scope adverbial fill the COMP position, while the sentence scope adverbial is attached to a higher S node:

84. In Ben's office he is a real dictator.



85. * In Ben's office he placed his new brass bed.



Since c-command may involve only one higher node of the same category, c-command prevents coreference only in (85).

If a language processor is to apply c-command, it must have as input surface phrase markers such as (84) and (85), or construct them from the incoming speech signal. Notice that in order to be able to determine which type of analysis is appropriate, the processor must

have access at least to the verb: for be, S-attachment is correct, while for place, VP-attachment is correct. This then determines the placement of COMP, which is the decisive factor in allowing or blocking coreference, according to Reinhart.

However, this seems an inefficient routine for a parser. A simpler solution would be to determine coreference options directly, by referring to the following interpretive rule:

86. A pronoun may not be coreferential with an NP in a preceding PP of sentence scope.

This obviates the need to work out the c-command domains. Why, then, does Reinhart choose the more complicated route?

The answer lies in Reinhart's contention that the COMP analysis is purely structural, and her desire to provide a unified model of coreference. In order to do so, she must demonstrate that the scope information is not in fact a semantic distinction by providing syntactic evidence that the two PP types are different. The evidence involves the following examples of Though Movement (87) and Pseudo-cleft Formation (88):

87. An absolute dictator though Ben is in his office...
*Place his new brass bed though Ben did in his office...

88. What Ben is in his office is an absolute dictator.
*What Ben did in his office is place his new brass bed.

These pairs show that each rule must operate on an entire VP; when the PP is extracted from the VP, ungrammatical sentences result. These can be remedied by maintaining the original VP structure:

89. Place his new brass bed in his office though Ben did, he...

90. What Ben did is place his new brass bed in his office.

This is a standard form of syntactic argumentation, but I do not

find it [illegible]ing; it is simply an attempt to save the analysis from splinte[illegible]eparate syntactic and semantic components. In fact, one migh[illegible] argue that the restrictions on Though Movement and [illegible]docle[illegible]mation are merely syntactic reflexes of the semantic se[illegible] dis[illegible]ction. In both cases, the claim can be understood in two rat[illegible]r [illegible]fferent ways. First, in Reinhart's view, it is COMP position that [illegible]rmines the scope interpretation, in accordance with the notion that [illegible]uctures are first generated by the grammar, then assigned semantic i[illegible]rpretations. Secondly, from the parser's standpoint, COMP positi[illegible]is merely a redundant coding of scope, which can be interpreted [illegible]rectly from (86). Reinhart is free to choose the more computationally costly route in order to have c-command interpret coreference in all cases. A parser, on the other hand, can omit the c-command analysis.

2.13 Resolution biases and processing strategies

It often comes as a surprise to the non-linguist that syntactic rules governing anaphora do not state which NP goes with which pronoun. That is, the syntactic constraints are not deterministic resolution principles: while some variant of command theory may specify which NP's may not be coreferential with a given pronoun, one cannot determine the antecedent actually intended by the speaker on the basis of syntax alone.

Given the interest in syntactic processing strategies sparked by Bever (1970) and Kimball (1973), it is therefore somewhat surprising that the issue of preferred antecedence has not been seriously con-

sidered in the linguistic or psycholinguistic literature. That such biases do exist can be surmised from the fact that sentences such as (91) and (92) are not normally perceived as ambiguous:

91. In the town where Doug decided to live with Rory, he is required to pay a special gun tax.

92. When John gave the money to Jeff, he glanced at George.

in spite of the fact that in each case the alternative NP may function as the antecedent:

93. In the town where Julie decided to live with Rory, he is required to pay a special gun tax.

94. When Sue gave the money to Jeff, he glanced at George.

The syntactic, discourse, and pragmatic basis of such biases has not, to my knowledge, been carried further than Jackendoff's (1972) mention of the subject resolution bias for reflexives such as:

95. The waitress showed Sheila a picture of herself.

(see page 12 above). In this section I shall sketch the kinds of variables that seem relevant to such research.

First, the relation of an NP to the verb appears to affect access to an ambiguous pronoun: subjects, as Jackendoff noted, are more available than non-subjects, even when the pronoun is in a higher sentence, as in (93) and (94). However, this is not always the case:

96. Fred gave a book by John's father to his friend.

97. Ellen was seen by Mary near her house.

98. Ellen saw Mary near her house.

99. 'Rosa won't like Mary anymore, with her mother interfering all the time.

100. Bill broke Dan's new glasses in his apartment.

101. Bob showed Bill the microphone that was hidden in his office.

In these cases, a genitive pronoun in a PP is equally likely to be coreferential with the preceding subject and with objects, genitive NP's modifying objects, and object agents. It seems too that any slight bias is attributable to pragmatic, rather than syntactic, variables: Resolution to Bob seems better for (101) if one considers it more likely that a person would reveal a bug in his own office to someone else, but resolution to Bill seems better if one thinks it more likely that a person would reveal a bugging to its victim.

In still other cases, the bias involves the extent to which a sentence is more naturally interpreted with internal resolution. The following two cases suggest that external resolution is sometimes preferable:

102. We'll just have to fire him, whether Smith knows about Jones or not.

103. We parked her car near Mary's house.

Conversely, sentences for which internal resolution is quite normal:

104. John said that he was sick.

are easily given external resolution by providing a topic NP:

105. George arrived today. John said that he was sick.

Judgments for a large variety of cases of free anaphora would provide an initial view of the strategies speakers use in interpreting these potential ambiguities; analysis of spontaneous speech samples would in all likelihood prove even more fruitful.

2.14 Anaphora and discourse structure

2.14.1 Anaphora without syntax

Despite the fact that anaphora is widely recognized to be a cen-

tral element of discourse cohesion (e.g., Halliday and Hasan, 1976), relatively little has been written about the effects of discourse structure on either anaphora resolution or constraints on intersentential coreference. There is, however, a school of thought which proposes to account for constraints on anaphora, as well as the relaxing of certain constraints, partly or entirely in terms of discourse structure.

Among the most radical proponents of this view are Keenan (1976) and Bolinger (1977,1979), whose general attitude is that there are no independent syntactic effects. Bolinger cites approvingly Keenan's comment that there are "no structurally statable restrictions", and Bily's (1977) claim that attempts at improving precede-and-command theory "are often mere ad hoc patchworks on the basic syntactic rule which does not work" (cited in Bolinger, 1979).

On closer examination, however, Bolinger's work is seen to have little relevance to the general issue of constraints on coreference; rather, it is directed toward the very different question of when in discourse speakers choose to use a pronoun over a lexical NP. Only occasionally does this affect the grammaticality of a given coreference reading.

Consider the case (first noted by Kuroda, 1968) of disallowed backward pronominalization for indefinite NP's as opposed to permissible coreference with definite NP's:

106. When he entered the room Mary saw (*a salesman) (John) smile.

In addition to pointing out that this is not statable as a syntactic constraint, Bolinger notes a context in which the condition is relaxed:

107(a) I guess I'm enough of a pushover to take any woman who

was reasonably willing. -- If she asked you, would you marry a widow with five children?

However, I believe that it is the conditional, and not the context, that relaxes the constraint:

107(b) *When/If she came to my door I told/would tell a prostitute to leave.

Note that substituting beggar for prostitute reveals further pragmatic effects. Furthermore, the contrast putatively exhibited in (106) was not found in the judgment data of Chapter 3 below.

Stronger claims about the role of discourse are advanced by Delisle (1973) (and later adopted by Kuno, 1975), who claims that free backward pronominalization, which later motivated the development of c-command theory, does not involve a rule of sentence grammar. His position is essentially that in sentences such as

108. The woman who is to marry him will visit Ralph tomorrow.

the pronoun refers not to the following Ralph, but to an antecedent in prior discourse. If this is true, it is unnecessary to account for backward anaphora at all. Instead, sentences such as (108) would be instances of forward discourse anaphora, and the full NP Ralph must be viewed as nothing more than a resurrection of the already-resolved antecedent.

For such sentences, Delisle claims that an antecedent must be established in prior discourse. There are several arguments against this view. First, Carden (1978) cites many counter-examples:

109. While he hadn't read the article, Associate Dean...Martin Gifford said...

Fully 51% of the instances of backward pronominalization in texts surveyed by Carden involved reference which could not be predicted

from prior context. Secondly, neutral contexts can be found in which intra-sentential resolution is obviously free and backward:

110.(a) Did you hear the news? (b) Did you hear Fred?
(c) The woman who is going to marry him will visit Ralph tomorrow.

And finally, the argument that discourse-forward anaphora excludes the possibility of backward is not only non-demonstrative, since both forward and backward anaphora might be involved, but also somewhat ludicrous when one then considers what coreference relation obtains between the pronoun and the later full NP: surely coreference is not blocked (as shown by 110a-c), but if it is not free, then it must be obligatory-forward, like reflexives. This cannot account for the ambiguity of 110b-c.

2.14.2 Kuno's discourse constraints

Kuno (1972 a,b; 1975) attempts to demonstrate that pronominal anaphora and reflexivization are "controlled primarily by nonsyntactic factors" (1975, p. 276). Here we shall restrict discussion to his comments on simple pronouns.

Unlike the majority of his contemporaries, Kuno's approach is transformational rather than interpretive; he assumes that pronominalization is a rule "which applies after all movement transformations apply" (p. 280), but views it as a replacement rule. Rather than specifying the conditions under which a pronoun and a lexical item may be coreferential, Kuno prefers to view the matter in terms of the discourse conditions under which a particular NP may be replaced by a pronoun. In this respect, his goals are similar to those of Bolinger.

Kuno's principles are divided between Discourse and Sentential factors. The most general of the discourse rules is the requirement that all NP's with referents which are "clear from the preceding discourse" (p. 280) are to be pronominalized.

Although Delisle (1973) made a similar assumption, it is not a necessary one. For example, reidentification is a frequent device, both in discourse and in single sentences:

111. Why would you give Oksana a pair of pink Pierre Cardin socks?
Because Oksana likes pink socks!

112. Whatever Lola wants, Lola gets.

In general, then, this constraint cannot explain coreference between lexical NP's.

Kuno's second principle is that if an NP has a discourse antecedent, all but one occurrence of the NP should be pronominalized, subject to a variety of conditions, including the precede-and-command constraint and the predictability requirement. Although this is at odds with Carden's demonstration that unpredictable backward pronominalization is also possible, Kuno's constraint might be viewed as useful for just those cases in which a sentence with free backward anaphora has no antecedent in prior discourse. The constraint prohibits pronominalization of the leftmost NP if its antecedent is not predictable:

113. Who did you calm when you saw him getting mad?
#I calmed him before Henry did something rash.
I calmed Henry before he did something rash.
(p. 283)

This implies that discourses with explicit mention of the antecedent are better; however, they seem to suffer from a lack of motivation

for reidentifying the second NP:

114. What did you do when you saw Henry getting mad?
?I calmed him before Henry did something rash.
I calmed him before he did something rash.

Moreover, the generality of the constraint is uncertain. I strongly disagree with Kuno's judgments in cases such as:

115. What should I do next?
*Go to your house, and if you can lift it, bring back the big rock...

116. What should I do next?
*If you can find him, ask John.

Not only are these acceptable pronominalizations, but they sometimes break Kuroda's indefiniteness constraint, despite Kuno's claim that accounting for indefiniteness effects is one of the advantages of his own approach:

117. Before I could stop it, a huge boulder rolled over the edge and crushed my new bike.

Furthermore, some of Kuno's own examples serve to demonstrate that for certain apparent violations of his constraint, sentence-internal coreference is indeed quite possible:

118. What do people eat in their lifetime?
By the time he reaches the age of 70, the average American consumes...

Perhaps Kuno's objection to this discourse stems from the fact that it uses a pronoun to shift the topic from people to the average American. In other words, there appears to exist a weak discourse constraint against using backward anaphora to introduce a new topic, since the topic pronoun will normally be taken to refer to the current topic. This is evident in (119), in which a topic shift, even if intended by the speaker, cannot be inferred by the listener:

119. Fred sold me a car last week. When he comes over this afternoon, tell John to give me a call.

Such a constraint is well motivated, but has little to do with Kuno's point; in his example, forward pronominalization is neither better nor worse:

120. What do people eat in their lifetime?
By the time the average American reaches the age of 70, he consumes...

Yet the fact that topic shift lies at the heart of the problem simply serves to demonstrate that people and the average American are not clearly related; in order to see the anaphoric relation, one must suppress the topic shift and infer that people and the average American are intended to be coreferential. To clarify this relation, one has only to substitute Americans for people in the original example:

121. What do Americans eat in a lifetime?
By the time he reaches the age of 70, the average American consumes...

Thus it would seem that Kuno's second principle is quite useless.

Another of Kuno's discourse constraints is the Constraint on Predictable Theme Pronominalization, which states that Pronominalization cannot apply when the to-be-pronominalized NP is both the theme of its sentence and predictable from prior context:

122. #If John can, he will go to see a movie.

123. If he can, John will go to see a movie.

In (123), for example, the subject of the matrix clause is thematic and predictable in that the prior question refers to John. However, Kuno's constraint seems to apply to the wrong NP: what is required is not a prohibition against a thematic pronoun, but one against a non-thematic full NP:

124. What will John do this Sunday?
If he can, he will go to see a movie.

In (124), the discourse breaks the predictable theme principle, but is much better than (122). While increasing the distance between the initial NP and the later pronoun improves (122):

125. If he can, that is, if Marsha doesn't make him paint the garage, John will see a movie.

no manipulation will make the full nominal natural in the subordinate clause of (122).

At any rate, these facts are better accounted for by a separate condition, also proposed by Kuno, which disallows pronominalization in the "known" portion of a backward pronominalization sentence:

126. Did you see John when he came to Boston?

(a) *Yes, I saw him when John came to Boston.
(b) Yes, I saw John when he came to Boston.
(c) Yes, I saw him when he came to Boston.

Notice, however, that it is more natural to use two pronouns, as in (126c).

It bears mentioning that (126a) is one of the disputed structures starred by Langacker but accepted by Reinhart (see page 22 above). Given the inappropriateness of this structure in a context where John is already "known", and the general principle that backward pronominalization will be interpreted as exophoric in discourse, it is not surprising that such sentences will often be judged ungrammatical. The psycholinguistic processes underlying such judgments are discussed in Chapters III and IV below.

To summarize, Kuno's arguments concerning the discourse basis of certain anaphoric processes are unconvincing, often based on questionable

judgments, and, although raising interesting questions about the role of discourse principles in anaphora, do not succeed in showing that anaphora is "controlled primarily by nonsyntactic factors" (p. 276).

2.15 Discourse cohesion effects: Relevance and topic shifts

In this section I consider the effect of discourse cohesion on anaphora interpretation. Consider these discourses:

127. Ann got home late last night.
Ethel says she was in an accident.

128. The doctor just arrived.
Gary is afraid he is having one of his attacks.

How can one explain the fact that anaphora is resolved to the topic of (127), but sentence-internally in (128)?

The answer seems to be that interpreting she as Ethel in (127) would result in an incoherent discourse, in which the speaker first mentions Ann's arrival time, then irrelevantly adds that Ethel had an accident. In (128), conversely, one infers that it is Gary who is having an attack, thus explaining the presence of the doctor. The alternative resolution, in which a doctor arrives and has an attack, is dismissed as unlikely.

Notice, however, that while resolution appears to be determined by a topic resolution strategy in (127), the more natural resolution in (128) involves a topic shift. Clearly, then, discourse topicalization is not the only extra-sentential factor governing resolution.

These examples demonstrate that anaphora itself is a vital component of topic maintenance and topic shift. Consider the following discourses, in which an error in interpreting the speaker's intended resolution also results in an error in topic interpretation:

129. Ann got home late last night. Ethel says she was in an accident, so that when she opened the door, she didn't notice that it was Ann standing there.

130. The doctor just arrived. Gary is afraid he is having one of his attacks, and that's why he forgot to call before coming over.

Even when one attempts to construct an incoherent discourse with potential inter-sentential anaphora, the reader attempts to find a means of determining either a reason for the topic shift, or an interpretation under which the topic is maintained:

131. Ann loves chocolate cake.
Ethel says she was in an accident.

It is difficult to read this discourse without speculating that Ann's love of chocolate cake somehow resulted from an accident, or that her love of chocolate cake led to the accident, or that Ann's love of chocolate cake resulted in Ethel's accident. The least likely interpretation is that the speaker has spontaneously uttered two completely unrelated sentences.

The mechanism underlying these observations is Grice's (1975) conversational maxim, "Be relevant". The search for relevance is an essential ground rule in co-operative conversation, and is crucial to the description of how shared knowledge affects the structure of linguistic interactions (Smith, 1982; Sperber and Wilson, 1982). In (128) and (130), the listener also makes use of world knowledge about doctors and house calls in order to establish coreference and topic structure. The modelling of such <u>scripts</u> is an important goal of both cognitive psychology and artificial intelligence (e.g. Schank and Abelson, 1977).

In summary, then, topic maintenance and topic shift are not simply factors which help to resolve anaphora; rather, they are often determined by how anaphora is resolved in a given context.

2.16 Scene shifts

Chafe (1976) has hypothesized that items leave a listener's consciousness when the conversational focus shifts from one scene to another, and that a co-operative speaker no longer refers to such items anaphorically (or with reduced stress associated with givenness) after a scene shift is realized. Webber (1979) provides this example:

> 132. Fred left his niece at home and headed for the zoo with Mary and John...When they got to the zoo,...the girl... (p. 1-7)

Here, the fact that the niece does not participate in the zoo scene means, in Chafe's terms, that Mary is the only feminine NP currently in the listener's consciousness, such that definite reference may be used unambiguously (the girl); in fact, it even allows the reader to infer that Mary is relatively young.

However, it is unlikely that this phenomenon has anything to do with the psychological notion of consciousness, which is normally taken to mean, roughly, that a concept is currently active in short term (working) memory. Chafe circularly defines consciousness as the domain of entities which can be referred to in this way, but it can readily be demonstrated that scene shifts do not necessarily eliminate NP's from the former scene as antecedents for a pronoun mentioned after a scene shift has taken place:

133. ... and that's why Bob had decided to stay behind in New York. Meanwhile, off in Cairo, Greg was writing to him every day.

If Bob were indeed no longer retrievable, the full NP would be necessary in the second sentence. Moreover, it is possible for the scene shift itself to invite the inference that the antecedent must be in a prior scene:

134. ...and that's why Bob had decided to stay behind while Fred and Greg taught Arabic at the American school in Cairo. So a month later in Cairo, Fred was writing to him every day.

In this case the pragmatics of the scene shift is unrelated to which scene is currently foregrounded, and the effect has nothing to do with consciousness, except in the circular sense adopted by Chafe.

The notion of scene shift as a discourse factor contributing to anaphora resolution is perhaps best viewed as only one of a very large number of inference types that can affect the process. Chafe's attempt to explain it in terms of memory processes, however, is unsuccessful.

2.17 Discourse referents and discourse model synthesis

Karttunen (1969, 1976) examined the conditions under which an indefinite NP establishes a "discourse referent", that is, an antecedent whose existence may be inferred from the very fact that an anaphoric expression is used. He pointed out that this notion explains why an NP cannot always be referred to anaphorically:

135. Bill doesn't have a car. *It is black.

Such constraints on pronominalization are reminiscent of Postal's (1969) "anaphoric islands":

136. Max is an orphan, but MY parents are alive.

137. *Max is an orphan, but MINE are alive.

where the implied NP the parents of Max can be contrasted with my parents, but not with the elliptical mine. Hankamer and Sag (1976) show that similar restrictions apply when referents are situationally rather than linguistically available:

138. (Hankamer attempts to stuff a 9-inch ball through a 6-inch hoop)
Sag: #It's not clear that you'll be able to.
It's not clear that you'll be able to do it.
(p. 392)

Webber (1979) approaches this issue in terms of what the discourse "makes available" (p. 1-15) as antecedents for pronouns: antecedents are very often not explicit NP's, but entities inferred from the ongoing construction of a discourse model. The following are only a few of the examples she provides:

139. Blend a cup of flour with some butter. Moisten it with some milk (it = the flour-butter mixture), then knead it (it = the flour, butter and milk mixture) into a ball.

140. A Rhodesian ridgeback bit me yesterday. They are really vicious. (they = the class of Rhodesian ridgebacks.)

141. John dunked Mary's braids in the inkwell. It made her cry. (it = the specific event of John dunking Mary's braids in the inkwell).

142. John dunked Mary's braids in the inkwell. It always makes her cry. (It = any occasion on which John dunks Mary's braids in the inkwell).

Notice in particular that the pronoun does not necessarily even agree morphologically with any overt NP (e.g., 140), and that the antecedent need not even be an NP (e.g., 141, 142).

Finally, Karttunen demonstrates that an antecedent may exist not [redacted]er's real world, but in a possible world created via [redacted]tional structure of certain verbs:

143. John regrets that he doesn't have a nicer car. He would keep it in a heated garage. / *He keeps it in a heated garage.

The notion of discourse model synthesis (Webber, 1981) as an explanation for the generation of non-overt antecedents is an important technical advance for both psycholinguistic and machine comprehension models.

2.18 Summary of formal treatments of anaphora

In the foregoing review of formal analyses of pronominal anaphora, I have emphasized the diverse nature of the phenomenon. Despite attempts to explain anaphora entirely in terms of resolution biases, discourse structure, pragmatic factors, and other higher-level effects, it seems fair to say that the original question, the principle or principles governing the distinction between free and blocked anaphora at the sentence level, remains unanswered. None of the sentence-level theories is completely satisfactory, and although there is general agreement that the distinction is related to depth of embedding, it is far from clear that all such distinctions can be subsumed under a single interpretive rule, or even that such a rule must make reference to syntactic structure alone.

In Chapter III we shall consider some of the processing factors which affect coreference judgments. As pointed out in 2.10, coreference judgments are far from clear, and this is probably the most important stumbling block to be overcome in answering the fundamental questions about the syntax and semantics of anaphora.

CHAPTER THREE

COREFERENCE JUDGMENTS, COMPREHENSION STRATEGIES, AND BLOCKED ANAPHORA

3.1 Introduction

In Chapter II we examined various accounts of the syntactic basis for native speakers' judgments of the acceptability of coreference, and the problem of inconsistent responses to the judgment task, both among linguists and among naive speakers, was mentioned in passing. In this chapter we shall consider some of the complications which arise when one attempts to account in detail for these unexpected deviations from the predictions of formal models.

These deviations can be attributed to several factors, not the least of which is the fact that in the anaphora literature, as in many other areas of linguistic investigation, linguists' judgments with respect to the crucial sentences are not representative of the intuitions of naive speakers. In addition, we shall see that the task of judging coreference possibilities is not a direct window on speakers' intuitions; rather, judgment presupposes comprehension, and in the case of the unusual sentences which are used to test the formal theories, a wide variety of processing strategies may be used - strategies which

do not necessarily reflect the speaker's internalized grammar. Evidence for the existence of such strategies will be presented in this chapter, while Chapter IV reports a judgment experiment which confirms the hypothesis that coreference judgments are heavily influenced by analogy and contextualization effects. Ultimately, it will be argued that judgment data, both from linguists and from non-linguists, cannot be accepted as a reliable indication of the mental representation of grammatical knowledge in this domain, and that only production data (such as text counts and analysis of recorded speech samples) can give proper insight into the nature of coreference constraints.

The structure of this chapter is as follows. In 3.2 we shall consider some of the variables which can affect a listener's decision as to whether or not coreference is possible in a particular sentence. In the course of this initial discussion, I report the results of a survey of a small number of native speakers which suggests that a surprising number of blocked cases are normally understood as allowing coreference. Careful analysis reveals plausible processing explanations for this result. In 3.3 we focus on methodological problems inherent in the gathering of coreference judgment data, with particular reference to a paper by Carden and Dieterich (1980). This raises the question of how "contextualization", both explicit and implicit, can influence judgments, and leads to the adoption of the experimental technique used in the study reported in Chapter IV. The last section, 3.5, attempts to place the results of this study into perspective with respect to the use of judgment and comprehension data in linguistic investigations.

3.2 Variables affecting coreference judgments

3.2.1 Assumed stress

Akmajian and Jackendoff(1970) provide several examples of how syntactic coreference options may be altered by changes in stress:

144. After he woke up, John went home.

145. *After he woke up, JOHN went home.

146. After HE woke up, John went home.

I do not find these examples particularly convincing: apart from the intended noncoreferential readings, the same stress pattern may also indicate contradiction under coreference:

147. Did you say that BILL went home after he woke up?
No, I said that after he woke up, JOHN went home.

148. Did you say that Lee went home after SHE woke up?
No, I said that after HE woke up, Lee went home.

Since this stress pattern allows both the coreferential and the noncoreferential readings, it must still be considered a case of free anaphora.

Nonetheless, there do exist cases in which a change in NP stress brings about a real change in coreference options. Consider (149):

149. They parked his car near Ian's house.

This is normally interpreted as blocked. However, some speakers are unwilling to state that coreference is impossible, and I believe that this is related to the fact that while stressing Ian results in definite blocking, stress on house makes coreference possible:

150. *They parked his car near IAN's house.

151. They parked his car near Ian's HOUSE.

It is important to note that if sentences are presented in written

form (e.g., 149), subjects are free to interpret the stress pattern in whatever way they see fit. This could result in a judgment of absolute non-coreference, despite the fact that the subject might accept the same sentence as anaphoric had he or she thought of the alternative stress pattern.

This seriously undermines the notion that the essential predictor of anaphoric relations is syntactic form, and argues for the more complex view that syntactic and suprasegmental features operate as independent but interacting parameters in coreference assignment. The example under consideration is particularly enlightening in this regard, since even appeal to "normal stress" (see Schmerling, 1974, for a critique of this notion) cannot resolve the issue. The "normal" stress for (149) is (150), with stress on the possessive lexical NP, and no possibility of coreference. The precede-and-command analysis predicts that coreference should be blocked in both cases, while c-command predicts that coreference should be possible for both. Neither theory can capture the fact that the coreference interpretation depends on the interaction between this syntactic form and stress. Since each theory is wrong for one of the stress versions, it is impossible to decide whether stress should be viewed as relaxing the blocking rule for (151), or as blocking otherwise free coreference in (150).

3.2.2 Length

Akmajian and Jackendoff also claim that as a sentence is lengthened, a coreferential reading often becomes easier to obtain:

152. *In John's apartment, he smokes pot.

153. In John's brand new apartment on Scrae Street, he smokes pot.

They attribute this to the fact that

> as the sentence becomes longer, the relative stress levels of the NP's and pronouns decrease with respect to the clause containing them. (p. 126)

Lakoff (1968), Reinhart (1981) and Wasow (1972) note similar effects.

While several speakers whom I surveyed agreed that (153) is somehow more grammatical than (152), none objected to coreference in the latter; two of the judges corrected the sentence to "John smokes pot in his apartment", but this stylistic alteration should be viewed not as rejection of the coreferential reading, but only as a judgment that this means of expressing coreference is marked.

Furthermore, there is no evidence at all to support the notion that the NP's concerned are somehow destressed. Even if careful acoustic analysis were to find evidence of destressing, it is not clear how an imperceptible stress difference could have the same effect on semantic interpretation as the contrastive stress differences in (144) - (151).

Reinhart (1981) attributes the effect not to stress, but to one's perception of the "relative importance" of the PP as the sentence is lengthened: the longer PP is less a locative verbal complement than an independent proposition, and the matrix clause becomes a statement about the PP.

Of course, Reinhart's suggestion is ad hoc. She does not provide further evidence to support the notion that length and "importance" are correlated, nor does her hypothesis make quantitative

predictions about this correlation: how many words or constituents are needed to bring about a change in coreference? For example, O'Grady (1982) shows that the addition of only one word can bring about a significant change in interpretation:

154. ?His mother likes John.

155. Only his mother likes John.

In addition to the lack of agreement among speakers as to whether (152) is blocked, none of these explanations accounts for the improvement in (157) over (156):

156. *John told her that Meryl was lazy.

157. One of the members of the Academy seems to have told her that Meryl may be up for an Oscar.

Reinhart's analysis fails here: there is no possible scope explanation. Akmajian and Jakendoff's explanation also fails, since the same sequence (told her that Meryl) may be pronounced with identical stress, rate and intonation, yet the difference in the acceptibility of coreference remains.

It would appear, then, that all we can say with certainty about length and coreference is that a longer version of a sentence sometimes allows coreference, while a shorter version does not. This is somewhat reminiscent of Ross's (1967) Heavy NP Constraint, in that it invites a processing solution (e.g. an unspecified memory constraint) which cannot be fully explained by appeal to a specific processing theory.

Such evidence poses a difficult problem for both grammatical and psycholinguistic accounts of anaphora: should either type of model block coreference for the particular dominance relation involved, and

relax the constraint under poorly defined length conditions? Or should a theory allow coreference (thereby vitiating current syntactic explanations) and block it only when the sentence is "too short"? All of the above-mentioned theorists have adopted the former solution, but the fact that at least some speakers accept coreference even with the shorter versions suggests that the latter approach may be a better one. It is also possible that speakers who accept coreference in the short versions do so on analogy to the longer versions, and that although they would always interpret such a structure as blocked during normal processing, the self-conscious metalinguistic task of coreference judgment encourages the use of such analogical processes.

It should be obvious, however, that appeals to native speaker intuitions cannot be operationalized in a judgment task. Perhaps the only way to avoid the circularity of this argument is to abandon the notion that speakers' judgments can reveal their grammatical knowledge, and accept only production data as reliable. One could accept the original formulation of the length effect if preposed PP's such as that of (152) are very rare or non-existent. Lack of production of such forms is much better evidence of the operation of a constraint than inconsistent responses from confused subjects to sentence types they would not use.

3.2.3 Perceived dialect variation

Subjects performing a judgment task may be influenced by their impression that a particular form is characteristic of a different

social or regional dialect. One of the subjects I surveyed based her coreference judgment for (152) on her impression that Yiddish-influenced dialects might allow such structures. However, she was uncertain as to whether coreference would indeed be possible in such cases, since she was not in fact very familiar with the dialect. Householder (1973) also mentions the possibility that grammaticality judgments can be influenced by the speaker's perceptions of other dialects.

3.2.4 Connectives

Bolinger (1979) claims that syntactic accounts are weakened by the fact that the type of connective used between clauses can alter coreference possibilities:

158. She could pass for my sister, though Joan isn't related to me at all.

159. *He had promised to help me, but John refused.

This argument is simply wrong. First, by the simple expedient of substituting but for though in (158) and (159), we see that coreference is unaffected by the connectives:

160. She could pass for my sister, but Joan isn't related to me at all.

161. *He had promised to help me, though John refused.

Secondly, the fact that coreference is blocked in (159) is hardly an argument that anaphora is blocked in the sense intended by other linguists. Coreference is not possible in this particular token, but the syntactic form does allow coreference in other cases. Lack of coreference between John and he can be attributed to pragmatic factors; under the coreferential reading, (159) is contradictory, and one thus assumes that coreference is not intended. Note that coreference can

readily be obtained if the contradiction is removed:

162. He had promised to help me, but John later refused.

Similarly, coreference can be blocked in (158) for reasons of a semantic and pragmatic nature:

163. *I think she could pass for my sister, though my late friend Joan said that my mother looks too old for that.

These cases are nothing more than specific instances of pragmatic biases in resolving free anaphora. In such cases, Bolinger's use of the asterisk deviates from the practice of other researchers, who define blocked anaphora solely on the basis of syntactic relations.

3.2.5 Pragmatic content

We have just seen that a sentence with syntactically free anaphora may be judged categorically nonanaphoric for reasons unrelated to syntax. This poses another methodological problem for those who claim that native speakers' judgments are an important source of data, in that naive speakers, unable to distinguish between syntactic and pragmatic effects, might "erroneously" judge such sentences as nonanaphoric, firmly denying the possibility of coreference. Variation in the transparency of the syntactically acceptable but pragmatically odd readings of the following examples suggests that interactions between syntax and pragmatics will be difficult to untangle:

164. I'll feed him his favourite dishes every night, whether Rod likes it or not.

165. Robert says that he died in his sleep last night.

166. When Ben went off to Africa to work on that telecommunications project, he stayed behind in Paris.

Although each of the above exhibits free anaphora by all syntactic accounts, the degree of garden pathing seems to be related to the strength of the pragmatic bias, which varies from sentence to sentence (see Hirst and Brill, 1980, for experimental evidence for this type of variation).

There is an interesting interaction between the length factor which relaxes the constraint on backward anaphora with complement clauses:

167. *I told her that Meryl was up for an Oscar.

168. One of the members of the Academy seems to have told her that Meryl is up for an Oscar.

and the pragmatic variation just noted. Although (168) allows coreference, one finds no garden pathing in the following pragmatically biased version:

169. *One of the members of the Academy seems to have told her that Meryl has left for Boston.

This suggests that despite the length factor, the reader is far from committed to the coreferential reading when the last four words are being processed. One might then speculate on why garden pathing does not occur: the coreferential reading for the longer version is marked, these structures are initially processed as non-anaphoric, and a final decision as to the correct reading is put off until all the contextual information has been processed. Such complex linguistic interactions promise to provide a wealth of hypotheses concerning the processing strategies speakers employ during comprehension.

3.2.6 Ungrammaticality due to factors other than coreference

One factor which clouds the linguistic arguments is that sentences which are purported to be nonanaphoric are sometimes starred in the literature for reasons having nothing to do with anaphora, although the asterisk is then interpreted as signalling blocked coreference. Consider Reinhart's judgments on the following, which she considers to be problematic for c-command:

170. *To him, I spoke in Ben's office.

171. *For him, Ben's neighbours gave a party.

Reinhart does not mark stress, but it is reasonable to assume that she intends these to have stress on the preposition, since they are presented in contrast to the following grammatical examples with stress on the prepositions:

172. Behind him, Ben keeps a gun.

173. Under him, Ben's mother found a mango.

I presented the starred sentences to five speakers, who were asked to decide whether coreference could obtain. All five commented on the strangeness of these sentences, presumably because of the fronted dative and benefactive NP's. When forced to make a decision, four of the judges said that coreference would be permissible despite this other strangeness. The fifth refused to speculate, arguing that if she had to imagine some new form of English in which such sentences were possible, she would not be able to predict how coreference might work under the new rules. In attempting to force coreference judgments on otherwise ungrammatical sentences, Reinhart has either been blinded by the fact that no interpretation is normal, or has greatly misjudged

the average person's ability to make semantic sense of asyntactic prose.

This brings to light the problem raised in 3.2.2 above: What do we make of linguists' and naive speakers' coreference judgments on bizarre sentences? Do such sentences really provide crucial information about the generality of an interpretive principle such as c-command? And who is to be trusted - the linguist, so strongly influenced by the perceptual set of his or her own predictions (see Spencer, 1973), or the naive speaker, who may opt out of syntactic processing entirely in order to make a necessary interpretation?

In this case, Reinhart seems willing to assume that a speaker can ignore the problem with PP position, and, like some linguistic automaton, apply the rules for normal cases to the new situation. While the application of a theory to new cases is a venerated scientific procedure, its application to non-cases is a departure from normal practice.

A better interpretation of Reinhart's asterisks is that PP fronting itself is disallowed under coreference for these cases. While this seems to be precisely the problem, it hinges on speakers' acceptance of coreference. Ironically, this is precisely what Reinhart's own theory predicts; had she noticed the true reason for the unacceptability of her starred sentences, she would not have considered the sentences to be problematic.

There is yet another enlightening problem with these examples. If one alters the stress pattern as follows:

174. To HIM, I spoke in Ben's office.

175. For HIM, Ben's neighbours gave a party.

the sentences appear - paradoxically - to be ambiguous between a free

and a blocked reading. That is, individual speakers find it difficult to decide whether the pronoun and the full NP may be coreferential. However, this is an insidious form of uncertainty, since the decision as to whether an anaphoric relation may exist should not be a question of degree (unlike other questions of grammaticality/acceptability), but a binary choice. Logically, either there does or there does not exist a reading under which coreference obtains. In other words, if a speaker decides that coreference is possible, even under very rare conditions, he should not be confused as to whether it might also be impossible.

The only plausible explanation for this kind of confusion is that the sentence is somehow ambiguous between a free and a blocked reading, and that the speaker sees each in turn, changing his judgment with each change in reading.

This is precisely what happens to these sentences when different discourse contexts are considered. In a context in which HIM is both contrastive and exophoric, no anaphoric reading is possible:

176. I hear you really embarrassed George when you criticized him.
Did you talk to Dave in front of the whole office too?
No. *To HIM I spoke in Ben's office.

In a context where the contrast cannot involve the prior NP, either the exophoric or the endophoric reading is possible:

177. Nobody even noticed when Mrs. Haynes moved away, but for HIM, Ben's neighbours threw a big party.

178. Sam says that nobody even noticed when Mrs. Haynes moved away, but for HIM, Ben's neighbours threw a big party.

Dinsmore (1979, 1981) has examined the role of implicit contextualization in grammaticality judgments, and argues that a theory derived

from judgment data will be heavily influenced by the (perhaps unconscious) appeal to both linguistic and non-linguistic contexts. This would explain the observed variation in intuitions about these sentences: some speakers should steadfastly deny the possibility of coreference; others should firmly accept it; still others, vaguely aware of the two readings, should waver between the two judgments.

Lastly, and to further complicate matters, analogy may also work across stress types, such that responses to (170) and (171) are based on the same kind of division of opinion found for (174) and (175). For example, Reinhart, who stars (170) - (171), may belong to the group which denies the possibility of coreference in (174) - (175), making implicit reference to the context of (176). My speakers, accepting (170) - (171), should also accept (174) - (175), if the basis for this judgment is the kind of context represented in (177) - (178), and provided that their appeal to context and analogy does not change over time.

In summary, detailed analysis of this example demonstrates that coreference judgments, both by linguists and by naive speakers, do not provide direct access to the speaker's grammar. Rather, they can be strongly influenced by processing principles related to comprehension and contextualization. Again, the extreme complexity of alternative explanations for grammaticality judgments, and the observed variation in speakers' response to the sentences, suggests that such data collection is an unreliable method for accessing speakers' intuitions concerning anaphora.

3.2.7 Analogy and depth of processing

Three of the four speakers in my initial survey allowed coreference on the following sentence, which was starred by Reinhart:

179. *In Ben's film about Rosa, she found a scratch.

One explanation for this judgment is that these speakers interpreted the sentence with the grammatical reading (not mentioned by Reinhart) under which Rosa finds a scratch in some unspecified object. However, all subjects denied having made this interpretation. A related explanation is that instead of perceiving the ambiguity and choosing the less common (but grammatical) reading, the speakers interpreted the sentence as coreferential on analogy with the much more frequent structures not involving picture nouns, and thus not allowing S-scope interpretation:

180. In Ben's letter to Rosa she found a typo.

Further support for the analogical basis of this judgment comes from the fact that even within the category of picture-noun sentences, ambiguities such as that involved in (179) constitute an extremely small range of pragmatic possibilities, since a picture noun will make the PP ambiguous between the locative and the VP-scope interpretations only if the last NP (scratch) can be interpreted as a physical part of the picture noun. Consider this example:

181. In Ben's film about Rosa, she found a husband.

Reinhart's judgment, although understandable when the sentence is carefully dissected, bears no relation to how a given speaker might resolve the anaphora. Given the extreme rarity of conditions under which a sentence such as (179) might be uttered, the speaker does not - or cannot - perform the required syntactic, semantic, pragmatic,

and contextual analysis to sufficient depth. That the confusion is attributable to processing complexity, and not to syntactic form, can be demonstrated by moving the PP to postverbal position:

182. *She found a scratch in Ben's film about Rosa.

This version of the sentence preserves the relevant dominance relations, but is much more readily judged non-anaphoric, since the adverbial is now more readily perceived as attached at VP.

Ultimately, however, one must question the claim that the c-command analysis reflects speakers' judgments, for it makes incorrect claims about the outcome of the judgment process. While it may indeed be true that speakers do not utter sentences such as (179) because of the constraints on coreference which they have internalized, one cannot claim that they will not interpret such sentences as coreferential. Certainly Reinhart did not gather naive speakers' intuitions on these sentences, although she certainly did consult other linguists. What her judgments reflect, however, is not speakers' judgments, but linguists' judgments of how speakers would judge the sentences if they only knew enough about syntax.

Nonetheless, this procedure may not be as far wrong as it may seem. If we consider that linguists base their opinions on an estimate of the likelihood that speakers would produce such a sentence with the coreferential interpretation, and that linguists are probably more aware of what people in fact say, then their opinions may be worth something after all. However, it is only to the extent that linguists are good estimators of production norms that their theorizing can be taken as a valid representation of linguistic representations.

3.2.8 Judgments on isolated sentences: Task demands

The basic premise underlying the use of grammaticality judgments in grammar construction is that native speakers (including linguists) are able to judge the absolute grammaticality of a sentence in isolation. With respect to blocked anaphora, however, this is a particularly risky venture. Even if a speaker would never utter a sentence such as those discussed in the preceding sections without intending external coreference, and even if s/he has never heard or read such a sentence without an appropriate context, it is not entirely clear that the generalizations governing this lack of production are available in the form of a blocking rule which would result in the rejection of anaphoric readings for such sentences presented in isolation. Given the task of judging whether coreference is possible, the speaker must consider the anaphoric reading and decide whether it is disallowed; this is a task of interpretation, and not of well-formedness *per se*.

Consider this analogy: given the sentence *John red book read*, a speaker can easily spot the ill-formedness. However, s/he can also interpret the string, and might well respond affirmatively to the question "Can this mean that John read a book?" When testing the acceptability of a given anaphoric interpretation, we also want to know two things: Is the sentence well-formed, *given a particular coreference reading*? Each of these aspects of the task is important, but each relates to a different aspect of the constraints on anaphora. By asking speakers to judge whether coreference is possible, we are really asking two separate questions; it may be impossible, in principle, to separate them.

3.2.9 Acceptability, preferred readings, and implicit context: More methodological pitfalls

In this section we examine in detail a paper by Carden and Dieterich (1980) which purports to provide a solution to the problem of contradictory judgments in the anaphora literature. However, I will argue first that their attempt to settle such controversies by experimental means is itself fraught with methodological problems, and secondly, that their interpretation of text count data - a technique which I favour - is flawed.

The paper in question deals with a disagreement over the coreference predictions for a sentence type which is a case of blocked backward anaphora according to precede-and-command (S-command), but which c-command would label as free:

183. We'll just have to fire him, whether McIntosh likes it or not.

Reinhart (1981) simply rejects Langacker's (1969) judgment of non-coreference for these structures, and suggests that it is the lack of discourse motivation for backward pronominalization, rather than a constraint on coreference, which accounts for the strangeness of the coreferential reading. Carden and Dieterich (hereafter, CD) suggest that the disagreement needs better empirical investigation before the matter can be resolved, and claim to provide an appropriate methodology for collecting native speaker judgments bearing on the problem.

CD maintain that linguists' judgments on such sentences are suspect, since they may be influenced by what Spencer (1973) has termed "perceptual set", arising from their familiarity with the theo-

retical issues involved. Specifically, Reinhart and Langacker, and their readers, may have been influenced by examples surrounding the crucial one. Langacker contrasted (184) with (185) to demonstrate noncoreference:

184. I knew Harvey when he was a little boy.

185. *I knew him when Harvey was a little boy.

while Reinhart contrasted (183) with (186), starred by both theories:

186. *He was fired, whether McIntosh liked it or not.

Notice as well that Reinhart's pair is intended to illustrate the different coreference options for subjects and non-subjects, one of the proclaimed advantages of c-command theory. If (185) and (186) are in fact found to allow coreference, then c-command loses one of its main claims to superiority. Other instances of possible difficulties with the subject-nonsubject contrast have already been mentioned (p. 27, (65)).

Several problems bear mentioning before we turn to CD's judgment data. First it is not at all clear that Reinhart's judgment on (186) is any less questionable than Langacker's opinion of (185). The fact that both theories predict noncoreference does not mean that this structure is "uncontroversial" to native speakers, who were not consulted on these judgments. This would have been but a minor quibble against CD, had they not later used (186) as a control item against which to measure relative grammaticality in their study. As it turns out, many subjects accepted coreference on this sentence as well!

Claiming that discourse coherence can outweight syntactic factors in grammaticality judgments, CD decided to present the target sentences

to naive speakers for judgment in a neutral context, and to ask them which NP, McIntosh (sentence-internal) or Smith (sentence-external) was the preferred antecedent. On a second pass with the same stimuli subjects were asked whether or not the other NP could be the antecedent.

The target sentences consisted of four types, modelled on the disputed (189) with the pronoun in main clause object position; the "uncontroversially bad" example, with the pronoun in main clause subject position (188); and two uncontroversially good examples, with the pronoun a subordinate clause subject and McIntosh subject or object of the main clause:

187. (CONTEXT) The directors discussed the situation with Smith all afternoon.
188. ("BAD") They finally decided that he would have to report to the new vice president, whether McIntosh liked it or not.
189. ("DISPUTED") They finally decided that they would have to put him under the new vice president, whether McIntosh liked it or not.
190. ("GOOD") They finally decided that, whether he liked it or not, McIntosh would have to report to the new vice president.
191. ("GOOD") They finally decided that, whether he liked it or not, they would have to put McIntosh under the new vice president.

The results showed that (188) patterns with (189), having fewer internal coreference judgments than either (190) or (191). CD therefore conclude that Reinhart was wrong, and that S-command is the better theory.

This conclusion is unwarranted on several grounds. First, it seems significant that both the crucial sentence and the "bad" one received a fair number of positive judgments (7 coreferential, 14 noncoreferential for 188; 7 coreferential, 19 noncoreferential for 189). Indeed, CD had found several instances of both types in a survey of

spontaneous texts, such as this "bad" one:

192. She was sitting with 8,000 people in Madison Square Garden at a New York Apples match, when Phyllis Rothstein of New Rochelle...

CD see their options as follows:

> We may have a good but (accidentally) rare construction; or we may have a bad construction occurring a few times because of errors.

Even in the face of the very large number of coreference judgments given by their subjects and the occurrence of perfectly acceptable examples in published materials, they maintain that since both theories predict (188) to be ungrammatical, acceptability judgments must be in error for both (188) and (189).

Notice, however, that no explanation for this large number of "errors" is provided, and CD do not consider the possibility that the sentences are rare in their samples because they are stylistically marked and characteristic of certain written registers. Perhaps if they had surveyed tabloids such as the National Enquirer or the Weekly World News, they would have found many more examples of this attempt to create suspense syntactically where none exists in the story itself.

On the other hand, if these sentences are perfectly grammatical, why did a majority of subjects not allow coreference? One reason is that by collecting relative grammaticality judgments in context, CD were asking in effect whether coreference was natural in [illegible] cular context they had chosen; this has no necessary bear[illegible]ether or not such sentences may allow coreference in other contexts, or in isolation.

Recall that I have provided several examples of how contextualization can bias anaphoric interpretation in free cases such that coreference is blocked in a particular token; this is not an argument that the sentence is syntactically blocked, and CD's evidence would therefore appear to be irrelevant, to the extent that the context they used did in fact bias interpretation in this way. In the case of their "neutral" context paragraph, it should be obvious that the foregrounding of Smith in the context sentence makes this NP more available as the antecedent for the main clause pronoun, in accordance with the principle that free backward anaphora is resolved externally when an appropriate NP is available in the preceding text.

This is, of course, a serious error in experimental design, but it had a rather interesting result: subjects more often resolved extra-sententially for (188) - (189) than for the other sentences. In these cases the pronoun is in the main clause. While preference for external resolution is irrelevant to the question of whether internal coreference is possible, it does raise the possibility that dominance structure is predictive of markedness with respect to internal resolution.

Solan (1978) reports some possibly relevant data. He presented children with four short discourses foregrounding a discourse topic NP (the cow), and followed these by four target sentences:

TOPIC: The cow...

193. The horse told the sheep that he would run around.

194. The horse hit the sheep after he ran around.

195. The horse hit the sheep with his stick.

196. The horse hit the sheep after his run.

His subjects, aged 5 to 7, chose the cow twice as frequently for (193) and (194) than for (195) and (196), although the most frequent responses by far involved internal resolution to horse or sheep. Solan attributes this to a processing strategy based on the fact that the pronoun is in an independent clause in the former cases, but his is a clausemate of horse and sheep in the latter cases. Unfortunately, his data analysis does not include internal coreference biases, which, as pointed out in 2.13 above, have been given far too little attention in the literature. If there is such a bias, however, Solan may be wrong in his general conclusion that children do not use context in resolving anaphora. It is quite likely that children would tend to resolve to the horse, viewing this as a topic shift within the short discourse. In a footnote, he reports that adult subjects choose the cow as the antecedent about 50% of the time, making no distinction between the two within-sentence conditions. Thus only the acquisition data support his notion that clausemate status affects coreference.

To return to the CD study, one might well ask why there were so many judgments of internal coreference, given the possibility of adopting a topic-resolution strategy. The anomaly can be explained by another design flaw: CD used one other passage (in a weak attempt to give their results greater generality), but in this case the context was truly neutral:

197. We looked all over for a suitable speaker. We talked first to John, then to Bruno, and finally to Doug.

In this case, CD succeed in having all three names equally back-

grounded (all are mentioned; none is the subject of the initial sentence). Their target sentences mention John and use the pronoun he in the same structures as (154). In this case we should expect subjects to respond with fewer external resolutions; unfortunately CD do not report their results separately for the two discourses. Nonetheless, it is likely that the large number of coreferential judgments can be attributed mainly to this condition, which offers the least biased test of the internal coreference question. Since overall, subjects accepted coreference on 40-50% of the "BAD" and "DISPUTED" cases, the figures for the neutral condition alone (i.e., 197) very probably exceed 50%. The fact that there is a significant difference in the number of 'yes' responses for the "GOOD" and the "BAD" or "DISPUTED" cases is hardly an argument against the possibility of coreference in the latter cases.

3.3 Conclusions: Judgments, comprehension processes, and language acquisition

In this chapter I have argued repeatedly that the collection of native speakers' judgments about sentences exhibiting putatively blocked anaphora is fraught with methodological dangers. Both non-syntactic linguistic factors (stress, length, dialect variation, semantic bias, ungrammaticality) and linguistic processing strategies (implicit contextualization, analogy, shallow processing) can render such judgments suspect, in that a judgment of possible coreference will not necessarily reflect the speaker's mental representation of the restrictions on anaphora. In other words, native speakers are not good judges of the boundary between free and blocked anaphora.

I have also argued that judgments on nonoccurring sentences should be discarded, and replaced by direct analysis of production data, since the method of collecting judgments is demonstrably not an unbiased procedure for accessing linguistic intuitions underlying performance. That I should suggest a return to the structuralist methodology of supporting grammatical descriptions with data from attested utterances may come as a shock to those for whom explaining linguistic intuitions is the central goal of linguistic theory. However, it is evident that linguists rarely consult native speakers at all, and that in practice, the intuitions described in the literature are only those of linguists. But does this mean that linguists' intuitions provide a better basis for theory construction in linguistics? Certainly not- the number of disputed cases in the experts' own writings, their failure to differentiate between syntactic and supra-segmental, pragmatic, or discourse-related linguistic factors, and their own susceptibility to processing effects, argues quite strongly against accepting their "expert" opinions as any more reliable than those of the native speaker. If linguists' intuitions are to be worth anything, they must reflect the constraints governing production of attested forms; if they do not, those forms themselves must determine the form of the grammar.

Spencer (1973) showed that non-linguists' grammaticality judgments are internally consistent, but different from those of linguists in 50% of the cases. I have claimed in the preceding two chapters that linguists' judgments for anaphora data are similarly non-representative, and that naive speakers should, given the processing explanations

set forth above, accept far more coreference than predicted by linguistic theory. It might be argued that, having accepted that process explanations suggest that judgment data are unreliable as a reflection of underlying competence, an investigation of speakers' reactions to blocked anaphora is pointless. However, there are at least two good reasons why this should be done. First, it is important to establish whether the prediction of very lax criteria for accepting coreference in the judgment task is indeed confirmed. Secondly, studies of the acquisition of constraints on anaphora rely on the assumption that the endpoint of the acquisition process is the adult grammar, as revealed either by linguistic analysis or by adults judgments on the crucial sentence. Since at an early stage children tend to accept coreference in a number of blocked constructions (e.g., Chomsky, 1969; Solan, 1978) this literature is geared toward testing for children's acquisition of the blocking principle. But given the arguments against using either type of judgment data as evidence for the structure of the adult grammar, it would be dangerous to use these as a point of comparison.

Taylor-Brown (1983), for example, reports that adult control responses to eight types of blocked anaphora ranged from 15% to 100% "blocked" responses, and she reports a mean of only 59% "correct" responses from adults. Noting that O'Grady's (1982) computational approach predicts a smaller number of such responses than c-command, she concludes that the former theory is better. However, O'Grady's theory is not a performance model; to the extent that subjects accept coreference because of processing strategies, their "free" responses

are inflated with respect to their internalized system, which is what O'Grady intends to model.

In addition to this general argument for a better understanding of adult judgment process, I would like to point out a glaring methodological error in certain studies which purport to demonstrate the child's acquisition of blocking principles. In many of these studies (e.g., Lust, 1977; Lust, Loveland and Kornet, 1980; Lust and Clifford, 1983; Solan, 1978; Tavakolian, 1977) the children are asked to act out a sentence exhibiting free anaphora when both a toy named in the sentence and another toy are available antecedents. Consider Tavakolian's examples:

198. To kiss the lion would make the duck happy.

199. For him to kiss the lion would make the duck happy.

200. That he kissed the lion would make the duck happy. (p. 176)

In each case, children tended to act out the kissing by moving an animal other than the duck, a result which was interpreted as demonstrating that they blocked coreference. However, this result has nothing to say about the acquisition of a blocking rule; rather, it shows again that for free anaphora, if the discourse or the situational context proved another possible antecedent, backward anaphora is interpreted as exophoric. The type of "blocking" revealed by such studies is not the type of constraint dealt with elsewhere in the literature.

In the next chapter we examine the claim that speakers accept far more coreference than predicted by popular theories of anaphora. If this is so, as predicted by the processing hypotheses laid out above, we

must acknowledge that linguists' judgments are not representative of those of other speakers, and further investigate the validity of the notion that even naive speakers' judgments provide clear access to the structure of their internalized grammars.

CHAPTER FOUR

JUDGMENT EXPERIMENT

Introduction

The process of deciding whether or not an NP may serve as the antecedent of given pronoun seems straightforward. Given a sentence such as (201):

201. Hé saw a snake in front of John.

one need only posit coreference, then attempt to find that reading in the sentence. This is essentially a semantic interpretation judgment, and is therefore qualitatively different from the task of judging the syntactic deviance of sentences whose interpretation is transparent:

202. ?That's the answer of which he got [redacted].

203. ?China is exciting for Sue and her friends to talk about visiting.

204. ?A picture will be published soon of the new prince.

Here, varying degrees of deviance make a yes/no judgment difficult, and this has been acknowledged in the extensive literature on the non-binary nature of grammaticality judgments (e.g., Bolinger, 1968;

Householder, 1973; Ziff, 1964).

The purpose of the judgment experiment reported here is to test the claim that in a judgment task, speakers will allow far more coreference than is predicted by our theories. The study was designed to reveal whether or not speakers have fairly direct access to the linguistic knowledge that seems to govern their production of pronoun-containing sentences. That is, I have assumed that many of the cases cited in the anaphora literature are in fact deviant, in that they would not be produced without a discourse antecedent. This amounts to a claim that linguists' judgments are simply an estimate of whether or not such sentences occur in spontaneous production.

A finding of rather liberal judgments would indicate that the judgment task is perhaps inappropriate for anaphora, since it is an interpretive phenomenon. It is likely that speakers' production constraints govern the mapping from intended interpretation to syntactic form, but not the reverse mapping.

Method

Subjects. The subjects were 30 native speakers of English, recruited through personal contact. They ranged in age from 18 to 50 years (M = 30.1) and had all completed at least a high school education (M = 3.8 years post-secondary). There were 17 females and 13 males, assigned at random to two groups. Subjects were not paid.

Materials. Thirty-three sentences claimed by linguists to exhibit blocked anaphora were selected from the literature and rewritten when necessary in order to avoid duplication of sentence

content. These are termed the Test sentences. Thirteen were blocked forward cases and 19 were blocked backward; one contained two pronouns. Included in this total were several structurally related pairs. For nine members of the Test set, and for three of the matched blocked pairs, an additional structurally related item exhibiting free anaphora was included in order to allow for comparisons between free and putatively blocked items. Thus, there were 33 blocked and 12 related free sentences.

There were also 30 distractor items, 15 clearly free cases (the YES items) and 15 clearly blocked (the NO's). In the latter category, mismatched gender cues guaranteed that a noncoreferential interpretation would be imposed. These items served as clear cases against which subjects could judge the Test items.

For each of the 75 sentences, one additional "Implication" item was written. This always asserted coreference between the pronoun and the crucial NP in the Test sentence. Thus, the implications were true for the YES items and false for the NO's. Their status with respect to the Test items is the measure of interest in the study.

Two lists were drawn up. On list I, the Implication preceded the Test sentence as in:

205. Implication: Donna is quite fond of her car.

206. Test: Fond of Donna's car though she is, I like it even more.

On list II the Test preceded the implication.

These sentence pairs were recorded in random order by a trained speaker, with approximately 7 seconds between pairs. During the recording of List II, the List I tape was used as a model, and

discrepancies between the two tapes in terms of wording, stress, speech rate and intonation were identified by a linguist, then corrected.

List I is presented in its entirety in Appendix A.

Procedure. Subjects were tested in groups of three to six in a quiet room. They were first read the appropriate instructions (see Appendix B), which require that they listen to each pair of sentences and (i) indicate whether the Test sentences contains the Implication by checking the appropriate space on the response sheets (Appendix C); (ii) rate their confidence in that decision on a scale of 1 (low confidence) to 5 (high confidence); and (iii) indicate whether that judgment reflects the manner in which they first interpreted the Test item by marking YES or NO on the response sheet. The instructions had been pretested on four subjects and rewritten until they were considered unambiguous and easy to follow.

Subjects were then allowed to ask questions. Questions concerning the decision process were answered with instructions to listen to the meaning of the Test sentence and to decide whether it could be construed as containing the meaning of the Implication.

Total running time, including instructions, was 35 minutes.

Results

Order of Test and Implication. The effect of the difference in presentation order for List I and List II was tested by means of three sets of 75 t-tests, one per item. First, none of the 75 items differed reliably in confidence scores, measured as the absolute value of the signed judgment-confidence score, from List I to List II ($p > .1$, two-tailed). Secondly, only four of the 75 items differed

reliably in signed confidence scores across lists: items 14, 37, and 40 had higher means on List II, while item 62 had a higher mean on List I ($p < .01$). Finally, items were rescored using +1 for all YES responses regardless of confidence rating, and -1 for all NO responses. Items 14, 37, and 40 had higher scores by this criterion on List II, while means for items 31 and 62 were higher in List I ($p < .01$).

A set of 75 Chi square analyses was run to compare the frequency of YES responses to each item across Lists. Only items 31, 37, 40, and 62 differed, in the direction already indicated for the t-tests. ($p < .01$).

An overall t-test comparing Test items for List I and List II also failed to reveal a list effect ($p > .1$). The three consistent differences (items 37, 40, and 62) in this analysis could be explained neither by post-hoc linguistic analysis nor by errors in writing and recording.

Error rates. There were fewer errors for NO items (12.2%) than for YES items (24.9%). If responses of "yes" to Test items are scored as errors, the error rate for that group is 54.9%.

Means for YES, NO and TEST items. An analysis of variance was performed on the mean scores per type, excluding the 12 free items which were matched with Test items. There were two levels of Order (implication first or second) nested under three levels of Type (YES, NO, and TEST). Only the Type factor was significant ($F(2,56) = 178.3$, $p < .001$). Table 1 presents the means for each condition. Differences among the means were evaluated by the Newman-Keuls procedure (Table 2). All means were significantly different from one another, with Test means falling between Yes and No means. The mean for Test items was

significantly closer to the mean for YES items (difference = 1.89) than to the mean for NO items (difference = 3.89).

	List I	List II	Combined Lists	Error rate
YES	2.40	2.32	2.36	24.9%
NO	-3.28	-3.55	-3.42	12.2%
TEST	0.25	0.70	0.47	54.9%

Table 1. Mean item scores and error rates, Experiment 1: combined judgment-confidence scores

TEST	YES	
3.89**	5.78**	NO
	1.89**	TEST

Table 2. Newman-Keuls test of differences among means, combined List scores, Experiment 1. Cutoff = 0.47 for 2 steps, 0.64 for 3 steps, $p = .01$.

Means without regard to confidence. The differences just noted could be caused by (i) different numbers of yes and no responses for each type; (ii) differences in confidence across types, or (c) differences in both coreference judgments and confidence. Therefore, a second ANOVA was run on scores disregarding confidence, by assigning a score of -1 to NO responses and +1 to YES responses. Results are shown in Tables 3 (means per type) and 4 (Newman-Keuls test). Types was significant ($F(2,28) = 153.4$, $p < .01$). Test items were closer to the YES items in terms of mean number of yes responses.

	List I	List II	Combined
YES	0.52	0.51	0.51
NO	-0.69	-0.80	-0.74
TEST	0.07	0.18	0.12

Table 3. Mean judgment scores, Experiment 1 (Yes = +1, No = -1).

TEST	YES	
0.86**	1.25**	NO
	0.39**	TEST

Table 4. Newman-Keuls test of differences among means, judgment scores, Experiment 1 (Cut-off = 0.196 for 2 steps; 0.223 for 3 steps, p=.01)

The third ANOVA, again identical in design, examined confidence ratings only: the absolute values of the mean judgment-confidence scores were submitted to analysis. Again, only Types was significant ($F(2,28 = 32.09$, $p < .01$). Confidence ratings were clearly lower for the Test items, but the difference in ratings for YES and NO items was only marginally significant, just meeting the cut-off criterion for $p = .05$. Results are shown in Tables 5 and 6.

	List I	List II	Combined
YES	4.28	4.20	4.24
NO	4.45	4.30	4.38
TEST	3.97	3.72	3.85

Table 5. Mean confidence scores, Experiment 1.

TEST	YES	
0.53**	0.14*	NO
	0.39**	TEST

Table 6. Newman-Keuls test, confidence ratings. (Cut-off = 0.18 for 2 steps, 0.21 for 3 steps, $p = .01$; and 0.14 for 3 steps, $p = .05$)

Reversals. A response of NO ("not my first impression") in the third column of the response sheet indicates that the subject initially understood the sentence as blocked, but realized that the other meaning was also possible. The mean number of such responses was recorded for each item (possible total = 30 per item); these scores were then submitted to a one-way analysis of variance, ignoring the Order factor. The Between factor was Types and the Within factor was Items: In this analysis the 12 Yes items matched with Test items were left in the Test group. Types was significant ($F(2,72) = 10.145$, $p < .0001$). Duncan's multiple range test shows that YES and NO items did not differ significantly in the number of reversals observed (1.80 and 1.13 per item, respectively), while the Test items had significantly more reversals per item than either of the other types (Table 7).

Subgroup 1	Subgroup 2
NO (1.13)	TEST (3.48)
YES (1.80)	

Table 7. Types arranged in homogeneous subgroups according to mean number of Reversals by Duncan's multiple range test, Experiment 1 ($p < .05$)

Table 8. Scores for Test items, Experiment 1

Item #	Sentence	Mean	%YES
1.	Sue met him when Harvey was a little boy.	0.67	60.0
2.	I told him that Fred couldn't come.	-3.25	13.3
3.	We told him and Sue that Jack was a fool.	-2.25	20.0
4.	One of the members of the Academy seems to have told her that Meryl is up for an Oscar.	-1.49	30.0
5.	After he woke up, ERNIE went home.	2.50	50.0
6.	The chauffeur parked his car near the banker's HOUSE.	-0.57	40.0
7.	How obnoxious to Ben's friends he is.	0.50	53.3
8.	For Irene's husband, she would give her life.	1.35	66.7
9.	Fond of Donna's car though she is, I like it even more.	1.00	60.0
10.	To him I spoke in Ben's office.	-2.17	23.3
11.	For her, Betty's neighbours threw a party.	3.40	90.0
12.	Beth showed him her new tricks in the lawyer's jacuzzi.	0.10	50.0
13.	He was hit on the head before the lecturer had a chance to say anything.	-1.94	33.3
14.	In Marilyn's jewelry box she put her valuable bracelet.	1.97	73.3
15.	For her new glasses, how much did she pay?	2.34	73.3
16.	She is riding a horse in Gary's picture of Jane.	0.97	66.7

Item #	Sentence	Mean	%YES
17.	It bothered her that Marie had failed.	-2.27	26.7
18.	~~The gossip~~ about him upset Rick.	1.70	70.0
19.	Near Dan, he saw a snake.	-2.72	33.3
20.	Beside her, Kate noticed a leprechaun.	2.88	83.3**
21.	When he entered the room, Mary saw a salesman smile.	-1.06	33.3
22.	When she left the office, Bob saw the receptionist smile.	-1.30	43.3**
23.	In John's apartment, he smokes pot.	2.17	80.0
24.	In Sylvia's new condo near the waterfront she lifts weights.	3.57	90.0**
25.	In Frank's picture of Evelyn, she found a scratch.	0.30	50.0
26.	In Ben's ~~film~~ about Rosa, she found a husband.	3.57	86.7**
27.	We fired him, since Bert's behaviour had become irrational.	4.45	100.0**
28.	They finally decided that he would have to report to the new vice-president, whether McIntosh liked it or not.	-0.36	46.7
29.	We realized that whether she liked it or not, Anna would have to report the incident to the police.	2.24	76.7**
30.	In Carter's home town, he spends his most pleasant hours.	1.83	70.0
31.	Believe it or not, he is still considered a genius in Reagan's home town.	0.37	46.7
32.	In Trudeau's riding, he is quite well respected.	2.25	73.3**
33.	I'm willing to pay him $2000 for Jack's car.	0.07	50.0

Item #	Sentence	Mean	% YES
34.	For Cathy's T.V. I'm willing to pay her at most $300.	3.24	86.7**
35.	Michael's problems, he won't talk about.	2.27	76.7
36.	You can't talk to him about Ben's kids.	-0.96	40.0
37.	Marie's work, you can't talk to her about.	1.77	76.7**
38.	In George's living room he placed his new brass bed.	1.37	66.7
39.	In Fred's office he is an absolute dictator.	3.00	86.7
40.	With Rosa's peacock feather, she tickles people.	0.60	63.3
41.	With Rosa's new job, she'll make a lot of money.	4.33	100.0**
42.	When Lorraine finishes school, she promised Jim to go to Paris.	1.47	63.3
43.	When he finishes this course, Bill promised his father to get a job.	3.53	80.0**
44.	In Betty's wedding picture, she hopes to look like a lady.	2.37	80.0
45.	In Fred's graduation picture, he looks like a gorilla.	274	80.0**

** Sentence predicted to permit coreference

Item analysis. Individual items were examined in two ways. First, Table 8 shows all 33 Test items, as well as the matched free items, along with the percentage of yes responses and mean signed confidence scores. Choosing 20% as an approximation to the error rate, given the error rate of 12.8% for NO items, it is seen that only one Test item failed to meet this conservative criterion.

Secondly, the scores on a number of Test items were compared with those for matched free items (see Table 8). T-tests were run on items 19 vs. 20, 21 vs. 22, 23 vs. 24, 33 vs. 34, 27 vs. 28, 38 vs. 39, 40 vs. 41, 42 vs. 43, and 44 vs. 45. Within each pair the first item is the Test case and the second is the matched free case. Three such pairs showed no significant difference between the scores (see Table 9).

Table 9. T-tests for paired comparisons, Test and matched free items.

Item	Mean	S.D.	t	df	Difference	p (2-tailed)
19	-1.38	3.87	-4.24	28	4.00	.01
20	2.62	3.31				
21	-1.07	3.85	-1.00	27	0.75	n.s.
22	-0.32	4.22				
23	2.17	3.38	-1.85	29	1.41	n.s.
24	3.58	2.62				
25	0.24	3.78	-4.87	28	3.35	.01
26	3.59	2.54				
33	0.07	3.99	-3.42	29	3.16	.01
34	3.23	3.18				
38	1.37	3.66	-2.19	29	1.63	.05
39	3.00	2.90				
40	0.60	3.94	-5.18	29	3.73	.01
41	4.33	1.88				
42	1.47	4.08	-1.78	29	1.66	n.s.
43	3.03	3.31				

Item	Mean	S.D.	t	df	Difference	p
44	2.37	3.51	-0.41	29	0.36	n.s.
45	2.73	3.41				

Table 9 (continued)

One-way ANOVA's were also run on the three matched triples. Within each set there were either one or two blocked items. Means were compared with the Newman-Keuls procedure. The results of this analysis are summarized in Table 10.

Item	Prediction	Mean	S.D.	Contrast	Difference	p
27	free	4.30	0.99	27-28	4.67	.01
28	blocked	-0.37	4.02	28-29	2.60	.01
29	free	2.23	3.46	27-29	2.07	.05
30	blocked	1.83	3.46	30-31	2.20	.05
31	blocked	-0.37	2.77	31-32	2.80	.01
32	free	2.43	3.59	30-32	0.60	n.s.
35	blocked	2.27	3.36	35-36	3.25	.01
36	blocked	-0.97	4.11	36-37	2.94	.01
37	free	1.77	3.62	35-37	0.50	n.s.

Table 10. Results of one-way ANOVA's for 3-element sets of syntactically related items, Experiment 1.

For one set (35,36,37) one of the blocked items received the highest score, and for another (30,31,32) a blocked item received a score which was not significantly lower than the score for the free item.

Discussion

Summary of response patterns. The results of this study fully support the claim that the 33 Test items are more likely to be judged as allowing coreference than one would expect on the basis of linguists'

published judgments (55%, versus 12.2% for the baseline set of blocked sentences). Only one sentence failed to surpass a conservative baseline error rate (20%). Mean yes responses were both consistently higher than for the clearly blocked cases, and consistently lower than for the clearly free cases, whether one considers yes/no scores alone or combined judgment-response scores. Subjects also had less confidence in their judgments on these items than in the controls. However, the mean confidence rating of 3.85 on a scale of 1 through 5 indicates that confidence alone cannot account for the discrepancy between linguists' and nonlinguists' judgments.

The 24.9% error rate for free items was surprising. It suggests that subjects often do not accept the interpretation imposed by the Implication sentence; they consider their first reaction, even in the free cases, to be the only possible, or perhaps the only reasonable interpretation. Thus, one might infer that the yes response to a Test item is also a low estimate of the subject's true ability to accept coreference on these items. On the other hand, subjects were more often able to change their minds about Test items: there were more than twice as many "reversals" on these items. This shows that a subject's initial reaction is often conservative, but open to reanalysis. The apparent unwillingness to accept an alternative reading for the truly free cases and the apparent willingness to change an initial reaction for the Test cases suggests that the two item types were treated rather differently by the subjects. Nonetheless, the possibility remains that the figure of 54.9% is affected by the initial interpretation, and is therefore a conservative estimate.

Individual items. The fact that the Test items were given a fairly large proportion of YES responses follows from the process explanations discussed in Chapter III. In this section I will summarize the relationship between those arguments and the Test item results. Table 8 provides a list of the relevant sentences. Since no tests of significance were performed across sentences which varied in content, parts of this discussion are merely suggestive of the processing effects. However, the overall pattern of results follows the predictions of Chapter III.

Item 1 is the sentence from Langacker (1969) which was challenged by Reinhart (1981). Sixty per cent of the subjects agreed with Reinhart, and 40% agreed with Langacker. This response pattern brings up the question of whether the 40% who block coreference here simply have trouble thinking of a context in which such a sentence might be used, while the 60% who allow coreference were able to supply such a context. If this is the case, then the no judgments are subject to revision: one could show the subject a context in which coreference is allowed, and change his or her mind. To the extent that such an argument is reasonable, one could also argue that any moderately large number of yes judgments on a sentence represents only the proportion of subjects who were able to see that it is indeed acceptable, and that other subjects could be convinced. By applying this principle to the Test cases, we could conclude that all of these items are in fact acceptable with the coreferential reading.

Only Test item 2 was clearly rejected. Subjects thus

did not allow coreference between the subject of a complement clause and an object pronoun in a higher sentence, as predicted by all theories. Item 3 shows that by expanding the pronoun NP with a conjunct, one obtains higher levels of acceptance (20%), and in item 4, where the crucial part of the sentence is held constant and the rest of the sentence is lengthened, coreference judgments rose to 30%. This result was predicted in 3.2.2 above.

Item 5 demonstrates that Akmajian and Jackendoff's (1970) stress example does not result in a clear blocking of coreference (50% yes). In 3.2.1 I demonstrated that such stress patterns are acceptable in contexts where the stress indicates contradiction under coreference. Only half the subjects appear to have been able to provide an implicit context in which this type of sentence is acceptable. This confirms my claim that implicit contextualization can vary from speaker to speaker, and thus can account for at least some variation in judgment data. It also suggests that subjects can sometimes be relatively conservative about allowing coreference for cases which should be given an anaphoric interpretation by the grammar.

Similarly, item 6 tests my claim that stressing the head NP of a PP structure can make coreference acceptable in this otherwise blocked case: 40% of subjects accepted that reading. This may be a conservative estimate, since the alternative NP, the chauffeur, is clearly the preferred antecedent; perhaps if a non-matching NP such as she were used, subjects would have been more likely to accept the coreference reading suggested here.

Sentences similar to 7, 8 and 9 are cited by Reinhart as evidence

against precede-and-command, since they show blocked forward pronominalization: the latter theory cannot account for noncoreference in cases where the pronoun follows the antecedent. While I agree that speakers are unlikely to use such sentences anaphorically, they do not seem to be categorically blocked. Indeed, they were accepted by 53.3%, 66% and 60% of subjects respectively.

Items 17 (26.6%), 33 (50%), and 19 (33.3%) are also considered to provide crucial evidence in favour of c-command. In the first two, the subject c-commands but follows the pronoun, while in the latter, he c-commands (but does not precede) the other NP.

Items 10 and 11 are the "problems" for c-command which I discussed in detail in 3.2.6. There I claimed that the ungrammaticality of these sentences is unrelated to anaphora, and that analogy to fronted locative PP's and to stress variants which are acceptable under specific discourse conditions could result in judgments of acceptability. Of my subjects, 23.3% accepted coreference on 10, and 90% allowed it for item 11.

Items 12, 35 and 36 figure prominently in Reinhart's argument that analyses which try to account for constraints on anaphoric interpretation at the level of deep structure are inadequate, since fronting the adverbial makes coreference possible for some (item 12) but not all (35 and 36). However, only 50% of subjects accepted 12, compared with 76.7% and 40% for 35 and 36 respectively.

Reinhart cites items similar to 13 (33.3%) and 28 (46.6%) in her attempt to demonstrate that only c-command can account for the different coreference options of subjects and non-subjects. Obviously, agreement

with the restriction is not unanimous.

Reinhart's discussion of the complex interplay between the syntax of COMP structures and preposed adverbials depends on the ungrammaticality of items 14 (73.3%), 15 (73.3%), 16 (73.3%), 25 (50%), 30 (70%), 31 (46.6%), 38 (66.7%), 40 (60.3%), 42 (63.3%) and 44 (80%). In 3.2.7 I presented several reasons for such results, including analogy to less unusual examples and inability to process the sentences deeply enough. This set of sentences presents perhaps the strongest argument for the notion that judgment data and production data reflect the distinction between well-formedness judgments and interpretive judgments. Subjects' inability to reject these sentences as non-anaphoric demonstrates the difficulties found in using judgment tasks.

Items 23 and 24 purport to demonstrate the length effect discussed in 3.2.2. Interestingly, the difference between the short (80%) and the long (90%) versions was not statistically significant ($\underline{M}$ = 2.17 and 3.57 respectively).

Finally, items 21 and 22 test Kuroda's (1968) definiteness constraint. Here too, the degree of acceptance of coreference (33.3% and 43.3% respectively) did not differ significantly. Instead, opinions were divided on both sentences.

General discussion. Carden and Dieterich (1980) provide a recent update on the issue of the status of unclear judgments. One solution has been proposed by Chomsky (1957): base the theory on the clear cases, then use that theory to determine the status of the unclear cases. Katz and Fodor (1963) go so far as to suggest that sentences

on which subjects disagree should be rejected as irrelevant to theory construction. Neither solution is satisfactory. First, if we were to base our theories only on clear cases, syntacticians might have exhausted the grammar long ago. Secondly, one would like to be able to distinguish between sentences for which there exists a restricted range of acceptable contexts, and those for which no context can render the sentence grammatical. Third, one would also like to be able to distinguish between non-existent sentences which give rise to significant numbers of positive judgments for the reasons outlined in Chapter III, and truly acceptable sentences. The present results suggest that the former category, at least for anaphora, may be a significant one. Finally, it would be beneficial if we were able to draw a distinction between judgments based on well-formedness criteria and those based on interpretive criteria. In this experiment, it was shown that the Test cases, whether accepted or rejected, received lower confidence ratings. This could indicate that subjects were basing their judgments on interpretive criteria, but were somewhat influenced by the relative ill-formedness of certain items. However, it is likely that for some items no illformedness was perceived. For example, the COMP / adverbial items had response profiles that were very close to the 75% yes observed on the YES items. For such items, it is very difficult to distinguish between application of a processing strategy and simple lack of constraint.

Carden and Dieterich propose a solution of "cross-methodology validation" (treating as resolved any issue for which text counts and speakers' judgments converge). One problem with their method, however,

is that it ignores statistically rare, but apparently perfectly well-formed sentences which do not match the predictions of the theory (they are rejected as errors). Without a method for distinguishing errors from acceptable but rare items, this is unsatisfactory. A second problem has to do with their method of gathering data. By showing subjects putatively blocked cases in discourse contexts that encourage extrasentential resolution, they biased their results away from possible yes judgments. In the present study, it was shown that even for perfectly clear cases of anaphora, subjects tend to stay with their original interpretations; this was reflected in a 25% error rate where the criterion is whether or not coreference could obtain, and where subjects are allowed to register the fact that a given judgment was not their first impression. By providing a definitive exophoric resolution context, Carden and Dieterich most likely increased this subject bias even further.

I have attempted to show that in grammaticality judgments involving anaphora, linguists tend to use production criteria based on the well-formedness of a mapping from coreference options onto syntactic forms. Naive speakers, on the other hand, tend to use interpretive criteria, mapping syntactic forms onto comprehension strategies. It is therefore not too surprising, with hindsight, to see that different judgments can arise.

It is therefore important to note that grammaticality judgment does not provide direct access to a speaker's competence, if by competence we mean some representation of linguistic generalizations that is neutral between production and comprehension. Rather, the judgment process is

itself a combination of comprehension ("What does this sentence mean?") and production ("Could I say that with this meaning in mind?"). For anaphora, we seem to have shown that speakers do not have access to the interpretive principles governing their productions during the judgment task.

The gathering of judgment data is an enterprise which has not been subjected to the kinds of careful controls which social scientists apply to other types of data collection. For example, Hill (1961) and Makay and Sleater (1960) simply instructed subjects to rate the grammaticality of sentences, without further explanation. Spencer (1973) instructed subjects to determine whether the sentences were "complete and well-formed". Stolz (1969) used the following instructions:

> Listen to each sentence carefully, but do not waste your energy searching for hidden or metaphoric interpretations. We are looking for your immediate and intuitive reaction to these materials and are notinterested in how ingenious you can be in contriving your own private readings for them (p. 210).

Carden and Dieterich (1980) asked their subjects to underline "the word to which the circled pronoun refers" based on "the most natural interpretation" for a first pass, but including "all the possible natural interpretations" on a second pass (p. 10). In this study, I instructed subjects to listen for an "implication" which asserted coreference between the NP's of interest. It is likely that there are as many further examples of instructions to informants as there are linguists who collect this type of data. It is also likely that variations in instructions will have a profound effect on the kinds of data

obtained. Such words as "well-formed", "possible", "natural", "intuitive reaction" and "implication" will tend to bias subjects in one way or another.

My interpretation of the anaphora literature leads me to conclude that linguists are more interested in the well-formedness aspect of anaphora than in the subject's ability to interpret a given pair of NP's as coreferential. Since this requires a judgment of well-formedness, it would be useful if we were to adopt certain conventions for the questioning of our informants, whether they be naive speakers, fellow linguists, or even ourselves. Specific instruction to judge whether or not the speaker would use such a sentence in any conceivable circumstances in order to convey the coreference meaning of interest should result in more interpretable data. It is also to be hoped that an agreement of this sort would obviate the need for the regular appearance of studies such as this, which after so many years of work within the transformational-generative tradition can still call into question the validity and reliability of the data upon which linguistic theories are built.

Some suggestions can also be offered regarding the refinement of the judgment experiment reported above. First, approximately 70% of the sentences which Reinhart analyzed with a COMP analysis in order to save the c-command analysis were judged acceptable by the subjects. This suggests that this category in particular should be examined in more detail. One improvement would be to retest subjects after a fairly long interval (say, six months) in order to determine the

relationship between firmness of judgments and consistency within speakers. If these sentences, or other types in the study, prove to be not only highly acceptable to speakers, but also consistent within speakers, the argument against this aspect of c-command analyses is strengthened. At the same time, sampling of a larger number of sentences with larger groups of subjects would allow a detailed analysis not only of item characteristics due to syntactic form, but also to individual differences in judgment strategies. For example, subject characteristics such as verbal I.Q., vocabulary size, reading ability, age, etc. might turn out to be correlated with judgment patterns. Hockett (1958) refers to an early article by Bloomfield entitles "Literate and illiterate speech" (American Speech, 1927, 2, 432-441) which suggests that certain aspects of linguistic competence might vary as a function of literacy. (See Appendix J for a preliminary analysis).

Perhaps there is a deeper issue involved here. Current research on learnability suggests that some aspects of syntax are difficult to infer from the linguistic data presented to a speaker, since negative evidence in the form of correction from another speaker and the ability to note what does not occur in the speech of others are not often available to the language learner. For example, the work of Lust, Loveland and Kornet cited above suggests that children accept coreference on sentences such as He said that Mickey was sick because there are many instances of backwards pronominalization which are free, and few opportunities for the child to learn that its

interpretation of a blocked case is erroneous. When one considers the kinds of sentences used to build linguistic theories of constraints on anaphora, it becomes evident that even adult speakers may be subject to such learnability constraints. It would be of interest to see learnability proofs for the aspects of c-command discussed in this chapter, for if that approach is a viable one, it should predict that linguists, with carefully sequences exposure to the rare sentences of interest, would have developed the appropriate constraints, while the linguistic laity would not. Such an approach would help to solve the problem of whether the acceptance of coreference by my subjects is a matter of the comprehension strategies outline above, or a matter of rule acquisition in the sense in which this term is used in the language acquisition literature.

CHAPTER FIVE

PSYCHOLINGUISTIC APPROACHES TO ANAPHORA RESOLUTION

5.1 Introduction

In the preceding chapters we have observed the extent to which syntacticians have been preoccupied with discovering structural properties of sentences which do and do not allow coreference. Psycholinguists, on the other hand, are interested in different aspects of the phenomenon. Garvey, Caramazza and Yates (1976) have described the division of duties as follows:

> It is considered the task of a linguistic grammar to assign a unique structural representation to any and all interpretations of a sentence, but the linguistic grammar is not required to account for a listener's selection of one or another possible meaning of an ambiguous surface structure. Rather, it is the task of a theory of comprehension to determine the means by which a listener assigns one or another possible readings to a given sentence token. (p. 239).

In Chapters II and III we examined difficulties in separating the two approaches, since the theory of comprehension will, in part, determine whether or not a sentence exhibiting pronominal anaphora is viewed by the speaker as ambiguous. In the following chapters I will examine aspects of comprehension which are more directly related

to how a reader or listener chooses an antecedent from a set of two or more candidate antecedents in the discourse.

Study of the cognitive processes underlying anaphora resolution is motivated by our awareness of the significance for general language comprehension models of an aspect of linguistic processing which cuts across three of our most sacred dichotomies: syntax vs. semantics, sentence vs. discourse, and linguistic vs. pragmatic knowledge. It is perhaps significant that while Springston (1975) was able to cite only one inconclusive study of adults' processing of pronouns, we now have fairly specific models of how the human processor exploits many kinds of cues in conducting antecedent searches through memory, and of how potential antecedents are selected, activated, and integrated into ongoing representations of the discourse.

In 5.2 I review the experimental literature on anaphora resolution. Research in this domain has covered an impressive range of topics over a relatively small number of years. These include:

(i) lexically-based resolution biases

(ii) the role of situational pragmatics

(iii) the processing consequences of choosing a pronoun over a lexical NP

(iv) memory scan models of the antecedent search

(v) the notion of "search domain"

(vi) activation effects associated with resolution.

Section 5.3 presents a brief overview of the more general question of whether language comprehension is essentially a top-down or bottom-up process, and a discussion of the relevance of anaphora research to this problem. The chapter ends with a summary of the questions

left unanswered by the experimental literature, and provides the general rationale for the experiments described in Chapters VI and VII.

5.2 Experimental approaches to anaphora resolution

5.2.1 Pronouns vs. lexical NP's: Effects of definiteness

Schlesinger (1968) presented a series of experiments in support of his argument that syntactic analysis is sometimes completed only after an attempt has been made at semantic interpretation based on comprehension strategies. One of these studies attempted to show that "sentences beginning with semantically indeterminate words" (p. 140) would be more difficult to process than those with determinate words. Such a result would suggest that very early in the sentence an attempt at interpretation would be hindered by the listener's inability to construct an on-going representation. The alternative hypothesis, one held implicitly by Katz and Fodor (1963) and processing models based on their theory, is that interpretation would be attempted only after a full syntactic representation had been worked out, at which time this sort of indeterminacy would have resolved itself.

Schlesinger used a stop watch to collect reading times for Hebrew sentences with forward and backward pronominalization, similar in form to the following English examples:

207. After he met the Chief of Staff, Mr. Churchill left for the House of Commons.

208. After Mr. Churchill met the Chief of Staff, he left for the house of Commons.

Readings times were indeed slower for (207), and although the effect did not reach statistical significance, it did hold for each item

tested. Reading times were even slower for:

209. After Mr. Churchill met the Chief of Staff, the Prime Minister left for the House of Commons.

If the first finding is indeed a real effect, one might conclude that it results from having to keep the pronoun unresolved in surface memory until an antecedent is found. Tanenhaus and Seidenberg (1981) found support for a similar hypothesis in a study of interclause relations and pronoun processing. They reasoned that a clause should function as a "processing unit", i.e., be integrated into the on-going semantic representation and lost from surface or working memory, only if a complete interpretation can be made. If the clause is available in working memory, probe latencies for words within it should be relatively fast, compared to clauses which have been recoded. For sentences with a pronoun preceding its antecedent, subjects were faster to respond to a word rhyming with a test word in the next clause. Consider this example:

210. When adults / they are cruel, kids often grow up to be maladjusted.

Given the test word bids, subjects responded more slowly to kids when the first clause contained adults, and thus had been recoded. Furthermore, this difference disappeared when the pronoun could be resolved to a sentence preceding the test sentence:

211. Some adults are extremely insensitive.

In the context of (211), the pronoun version of (210) can be resolved immediately, and this accounts for the increased latencies observed with such contexts.

Schlesinger's second finding may be related to Lesgold's (1972)

later demonstration that sentences with pronouns are better recalled than sentences with repeated nouns in a pronominalizable position:

212. The blacksmith was skilled and he pounded the anvil.

213. The blacksmith was skilled and the blacksmith pounded the anvil.

Not only was prompted recall better in the pronominalized versions, but more words were recalled from the clause not containing the prompt word. This implies that sentences are better integrated if they use a pronoun where "resurrection" of the full lexical item is unnecessary. Schlesinger's result shows that even when the use of an anaphoric definite NP has a discourse function, recognition of the anaphoric relation between two lexical items takes more time than pronominal anaphora resolution.

Without discourse motivation, a definite NP will normally be interpreted as noncoreferential with a preceding definite NP:

214. When Churchill left, the young lady was exhausted.

In Schlesinger's materials, inferring coreference is similar to the process of recognizing coreference for epithets:

215. When Churchill left, the old [redacted]ezer was exhausted.

In both cases the basis for the inference is world knowledge. Coreference would be blocked for (215) in the context of a description of the young Churchill, just as it might escape detection in (209) for a reader who has never heard of Churchill.

Yekovich and Walker (1978) have further shown that the interpretation of an indefinite NP such as a puppy:

216. The little boy played with a puppy. The puppy was only six weeks old.

is easier when the indefiniteness functions to signal reference to a specific individual not previously mentioned than if reference is to a nonspecific individual:

217. The little boy longed for a puppy. The puppy was only six weeks old.

This distinction is controlled by the verb (play vs. long for), and was termed directed vs. nondirected action by Yekovitch and Walker. Reading times for the second sentence were longer for items such as (217).

While the effect is undoubtedly real, it should be noted that the second type of discourse violates the discourse model assumptions proposed by Karttunen (1969), Postal (1969), Hankamer and Sag (1976), Webber (1981) and others (see 2.17 above). Such discourses are in fact incoherent and would be disallowed by a detailed model of English discourse.

It is also of interest to note that the distinction between directed and non-directed action is only one aspect of the discourse principles which govern the specific - nonspecific contrast. In French, for example, this contrast can be made explicit in a relative clause by use of the subjunctive:

218. Je cherche un homme qui conduise une belle voiture.

219. Je cherche un homme qui conduit une belle voiture.

while in English the difference, termed referential vs. non-referential is not overtly marked:

220. I'm looking for a man who drives a nice car.

Misuse of a nonspecific NP can also be spotted in discourses where

the use of one vs. him (or another personal pronoun) marks specificity; the discourse model assumptions in these cases are a matter of world knowledge as well:

221. I'm looking for a man with a good mind. I've always wanted to meet #him/one.

222. I'm looking for a man with a red hat. I've always wanted to meet him/#one.

The garden pathing in (221) and (222) results from pragmatic constraints on the likelihood that the indefinite NP in the first sentence is also a case of specific reference.

Thus, the constraint studied by Yekovitch and Walker is more general than their proposed verb feature.

5.2.2 Sentence-level interpretive biases: Verb-based effects

Springston (1975) investigated several constraints on coreference interpretation, some of which are similar to the verb feature analysis of Yekovitch and Walker (1978). For example, verbs such as like, which mark their subject as the experiencer of that emotion, bias the interpretation of embedded pronoun subjects to the subject of the higher sentence. However, no such bias obtains with pronoun objects:

223. Bill told Harry that he liked John.

224. Bill told Harry that John liked him.

The converse is true of verbs like bore, which mark their object as experiencer and bias embedded pronoun object, but not subject interpretation to the higher subject:

225. Bill told Harry that John bored him.

226. Bill told Harry that he bored John.

There is a pragmatic basis to the constraint: under the nonpreferred reading of the like and bore sentences, Bill performs the bizarre speech act of informing Harry of Harry's own reactions to other people. However, in contexts where this is a plausible speech act, the bias shifts:

227. Harry was having a long session with his analyst, Bill, about his relationship with his brother John. One of the most stunning insights came when Bill told Harry that John bored him.

Nonetheless, the locus of the bias is the verb, and there is nothing to suggest that considering it a lexical bias affects the main conclusions of Springston's study.

In a reading time study, Springston found that subjects comprehend sentences which are congruent with this constraint faster than those which are not. Thus, even for sentences in which gender cues alone would be sufficient to determine the antecedent, comprehension was faster for the biased cases:

228. John told Mary that he bored Sue. (unbiased)

229. John told Mary that Sue bored him. (biased)

Furthermore, violations of the constraint took longer to process, even when cues were present:

230. John told Mary that Fred bored her.

Each of these results can be interpreted in terms of on-line access to the constraint as the verb is being processed. In (229), resolution to John is predictable even before the pronoun is heard or read, and in (230) this prediction is not met by the pronoun. This provides further evidence that resolution is affected by prior context, as already suggested by Schlesinger (1968) and Tanenhaus and Seidenberg (1981).

Garvey and Caramazza (1974) described a third verb feature, implicit causality, which biases anaphora interpretation in because clauses:

231. George telephoned Walter because he wanted sympathy.

232. George criticized Walter because he wanted sympathy.

Verbs such as telephone seem to set up an expectation that if a cause is offered for the action, it will involve another event in which the subject (e.g., George) is the agent; verbs such as criticize, on the other hand, imply that their object (e.g., Walter) is the agent of the causal clause. As with the experiencer constraint, the effect is pragmatic, and is rooted in the knowledge of cultural norms governing the acts of telephoning and criticizing, combined with our attitudes toward people who seek sympathy.

Garvey, Caramazza and Yates (1976) showed that naive speakers are strongly influenced by this bias. When asked to complete spoken sentence fragments such as

233. The father scolded his son because he.....

both adults and 10 - 13 year old subjects were oblivious to the potential ambiguity, and were able to offer completions without difficulty. These completions show that the bias covers a continuous range from verbs which received an equal number of continuations consistent with each of the interpretations of the pronoun, to strong biases toward either the matrix subject or the matrix object. This held both within sentences, as above, and across question-answer pairs such as

234. Why did the mother punish the daughter? Because she...

Other factors had a modulating effect on the implicit causality strategy. Passivized versions of biased sentences showed a general

increase in the number of surface subject choices. For example, the following had more Walter interpretations than did (231):

235. Walter was telephoned by George because he wanted sympathy.

Verb negation, which has a complex effect on causal relations between clauses, also results in a stronger subject bias:

236. George killed Fred because he.....

237. George didn't kill Fred because he.....

Finally, the social status of the people referred to in the sentences interacted with the lexical and syntactic biases:

238. The assistant argued with his boss because he.....

239. The boss argued with his assistant because he.....

In (239) there is a weak bias toward resolution to assistant, contrary to the results for (238). Apparently resolution here is partly determined by the subject's interpretation of who is more likely to begin such an argument, and who is likely to be blamed for it.

Caramazza, Grober, Garvey and Yates (1977), like Springston, found that violation of implicit causality is associated with slower reading times. A sentence constitutes a violation of this constraint if its causality bias is different from the bias of its because clause:

240. John telephoned Bill because he withheld some information.

Slower reading times for these sentences would indicate that the bias takes effect during processing of the verb and the conjunction, thereby leading to a garden path effect when the verb phrase of the lower sentence is interpreted. This result was also found when gender cues were sufficient for resolution:

241. John telephoned Sue because she withheld some information.

Under the assumptions of the on-line resolution model being developed here, this result is not surprising. Whether the gender of the two NP's is different or not, on-line interpretation of the sentence leads to a prediction that the word after because will be a pronoun referring to the NP that is consistent with the bias to that point. We can also predict the precise locus of the garden-pathing for (240) and for (241): in (240) the violation of implicit causality cannot be perceived until the second clause has been integrated, while in (241) the violation can be perceived as soon as the pronoun is encountered.

However, Caramazza et al. conclude that both interpretations of the pronoun are considered during processing of the because clause in both types of sentences. This is viewed as support for the position of MacKay (1970) and Olson and MacKay (1974) with regard to the processing of other types of ambiguity, and is compatible with the Clausal Processing Hypothesis (Fodor, Bever, and Garrett, 1974), for they also claim that the final interpretation is chosen only at the final clause boundary. It is also inconsistent with my contention that resolution occurs on-line for (241), a position held by the on-line interactive (top-down) school of thought (e.g., Tyler and Marslen-Wilson, 1980). Such issues are examined in detail in 5.3.

The work of Tanenhaus and Seidenberg (1981) suggests that the claims of Caramazza et al. are wrong. In sentences where prior context provided sufficient basis for resolution, Tanenhaus and Seidenberg found evidence that the prior clause had been recoded, and hence the pronoun had been resolved. A similar method could be applied in this case: by probing for a word from the first clause, one could

examine probe latencies for two positions during the second clause and after the entire sentence has been processed:

243. John telephoned Bill because he***

244. John telephoned Bill because he withheld some information.***

245. John telephoned Sue because she***

246. John telephoned Sue because she withheld some information.***

*** Probe JOHN

The on-line interactive theory predicts that probe latencies will be longer for the unambiguous fragment (245) than for the ambiguous fragment (246), while the Clausal Processing Hypothesis predicts that latencies should be identical. Both theories predict identical latencies for (244) and (246); comparison to these positions would provide a baseline measure.

In a fourth study, Grober, Beardsley and Caramazza (1978) showed further effects due to the particular modal verbs used: "strong" modals such as must and ought to express necessity; "weak" modals such as can and may express possibility. Grober et al. hypothesized that weak modals would result in attenuation of the causality effect, while strong modals would augment it:

247. Nancy scolded Marge because she...

248. Nancy may scold Marge because she...

249. Nancy ought to scold Marge because she...

These predictions were borne out in a sentence completion task. Grober et al. also expected than when casality effects were relaxed, subjects would choose an antecedent in accordance with Sheldon's (1974) Parallel Function Strategy, by which a lower subject pronoun is under-

stood as coreferential with the higher subject. This result obtained for both the weak modals and for sentences containing non-causal but clauses:

250. John criticized Bill, but he.....

Caramazza and Gupta (1979) also examined interactions among implicit causality, passivization, and clause fronting in a timed antecedent naming task:

251. John scolded Bill because he was annoyed.

252. Because he was annoyed, John scolded Bill.

253. Bill was scolded by John because he was annoyed.

254. Because he was annoyed, Bill was scolded by John.

Clause fronting alone did not eliminate the implicit causality bias, but both passivization and a combination of the two conditions reduced the effect. The authors attribute this to two factors: the greater processing difficulty of backward anaphora, and the importance of topicalization in this resolution strategy.

Ehrlich (1980) has shown that implicit causality is best viewed as a general pragmatic factor rather than a lexical feature of the verb. She had subjects read sentences similar to those used by Caramazza et al. (1977) and name the antecedent. However, she varied the relation between the clauses in the materials by using three different conjunctions: because, and, and but:

255. Steve blamed Frank/Sue because he spilt the coffee.

256. Steve blamed Frank/Sue, and he spilt the coffee.

257. Steve blamed Frank/Sue but he spilt the coffee.

The effect was replicated for because and but, but not for the and

sentences. Ehrlich therefore argues that implicit causality reflects an interaction between the semantics of the verb and event relations as encoded in the conjunctions. This does not seem to be a serious criticism, since implicit causality is claimed to be a verb feature that affects the interpretation of causal sentences only.

Ehrlich also found that in sentences with a gender cue (the Sue versions of (255 - 257), changing the conjunction had no effect on reading times. This seems to argue that gender has priority over implicit causality in anaphora resolution. However, she does not discuss this result in relation to Caramazza et al.'s finding that the implicit causality effect is present even when gender cues disambiguate anaphora. Why should one effect obtain in the presence of gender cues, and the other be suppressed?

This seemingly anomalous result can be explained with reference to the two process models described above, and is in fact predicted by the on-line interactive theory. Notice that even with the more difficult (less coherent) sentences:

258. Steve blamed Sue and she spilt the coffee.

a subject would be able to state the antecedent without fully comprehending the relationship between the clauses, if and only if resolution were achieved on-line. However, when gender cues are absent, as in the first part of her experiment, the subject must look elsewhere for information which would allow resolution, namely the type of relation which holds between the clauses. The Clausal Processing assumption, however, makes the wrong prediction for this part of the experiment. If resolution were delayed until the clause boundary were reached, one would

expect to find equal facilitation due to gender, in addition to the effect of clausal relations, and therefore faster but similar latencies to those found in the first condition.

5.2.3 Intersentential interpretive biases

Hirst and Brill (1980) attempt to account for speakers' choice of antecedents from prior sentences in terms of the likelihood that the event described in the pronoun-containing sentence would involve the person named in the main or subordinate clause of a preceding sentence:

259. John stood watching while Henry fell down some stairs. He ran for a doctor / thought of the future / cried out in pain.

Here the pronoun of the first continuation is pragmatically biased toward coreference with John, the second is neutral, and the third is biased toward Henry.

First, Hirst and Brill collected speakers' ratings of the plausibility of each alternative, e.g.:

260. If John had stood watching, he might have run for a doctor.

261. If Henry had fallen down some stairs, he might have run for a doctor.

"Plausibility differences" were obtained by subtracting the ratings for one alternative from the other. These were then shown to correlate strongly with reading times for test discourses, as reflected in an antecedent choice task. Large differences (e.g., for the biased first and third continuations of 259) were associated with longer latencies, as subjects seemingly puzzled over the plausibility of each.

A more surprising result is that even when syntactic cues which

block coreference seem to be available on-line, the plausibility effect still held:

262. John stood watching. He ran for a doctor / thought of the future after Henry fell down some stairs.

Resolution should be possible, under the assumption of interactive processing, when the pronoun is reached, since the subject position allows resolution to a preceding NP. At the very least, it should be possible to discount later resolution by the time the conjunction after has been perceived, since coreference in the lower clause is blocked on syntactic grounds.

Hirst and Brill interpret this finding as evidence that the integration of the content of the two sentences is triggered by the pronoun, [illegible] resolution occurs only when integration of the second clause [illegible]y.

[illegible]s is counter-intuitive: one has neither an impression of un[illegible]inty of reference in the first clause when the second is being pro[illegible]sed, nor an impression of ambiguity resolution at the end of the sen[illegible]ce. Furthermore, we shall see in subsequent sections that eye-movement research (e.g., Ehrlich and Rayner, 1983) has shown that resolution is often completed during the fixation on which the pronoun perceived. Moreover, even bottom-up models of comprehension (e.g., Forster, 1978) assert that interpretation occurs at the next clause boundary whenever contextual information from preceding clauses can be used. In this case, the pragmatic information comes in the clause following the pronoun, well after sufficient information is available for resolution.

Hirst and Brill's result is most probably an artifact of their experimental technique. Their design leaves open two alternative interpretations. First, rather than measure global reading time for the sentences, they had subjects push one of the two buttons marked John and Henry for each trial in order to indicate antecedent choice. This is essentially a choice reaction time task, and as such requires verification of two alternatives after resolution has taken place. After initial processing is complete, subjects must consider each NP, even if they had not done so during resolution. Large plausibility differences may well facilitate this postprocessing decision, but it is unlikely that they affect resolution on-line.

A second flaw in the study is the fact that plausibility differences alone could account for the initial reading time differences, whether they occur during initial processing or during the antecedent choice task. Consider this example:

263. John stood watching. Fred thought of the future / cried out in pain after Henry fell down some stairs.

It is quite likely that the same pattern of response times would appear with no pronoun, simply because it is difficult to understand why a person would think of the future when witnessing an accident.

Since these design problems bias both reading times and verification times in the same direction, the counter-intuitive and otherwise uninterpretable result should not be taken seriously.

Lockman and Klappholz (1981) have speculated that in semantically neutral contexts an ambiguous pronoun will be interpreted as coreferential with a preceding main clause NP in preference to a subordinate clause NP, regardless of recency:

264. John cut the grass while Bill washed the car. He did a lousy job.

265. While Bill washed the car, John cut the grass. He did a lousy job.

Clark and Sengul (1979) did test explicitly for main/subordinate clause effects in similar materials, but found none. However, their materials were not controlled for plausibility, a factor which Lockman and Klappholz suggest will have a powerful effect on resolution times.

Carpenter and Just (1978) found that as they increased the number of sentences between the first and second mention of a definite NP, their subjects took longer to comprehend the second sentence:

266. The ballerina captivated a musician in the orchestra during the performance. It was the conductor who arranged the choreography. (...) The one who the ballerina captivated was the trombonist.

They attribute this in part to the fact that the second sentence constitutes a topic shift, which becomes more firmly entrenched as subsequent sentences are added. Apparently, definite NP anaphora is more difficult when the antecedent is no longer foregrounded.

Frederiksen (1981) investigated the effects of a number of intersentential variables on processing ease for sentences which were semantically constrained to make the intended referent clear. Hypotheses were then advanced concerning the ease of resolution under various other manipulations.

Subjects controlled the display of sentences, one at a time, on a CRT screen, and their reading times were recorded. At unpredictable intervals, a button press would visually emphasize a pronoun in the current sentence rather than display a new sentence, and subjects' vocal reaction times to naming the antecedent were measured by

a voice-operated timer. Frederiksen observed several effects that have already been discussed, as well as faster reading times when a target sentence maintained the previous topic rather than shifting to an earlier topic; when intervening sentences made no topic reference, no reduction in reading times was observed. Furthermore, reading and naming latencies were faster when the semantically constrained antecedent appeared in subject position than when it was a non-subject in the preceding sentence.

Another of Frederiksen's observations is that resolution times increased as the number of potential antecedents in prior discourse was increased (see also Kaplan, 1981). He claims that this supports a "reinstatement" theory in which all nearby NP's are automatically canvassed. This can be interpreted as a claim that resolution occurs on-line, since pragmatic information after the pronoun did not facilitate resolution:

267. The nineteenth century was (a period in which numerous immigrants came to America) (a period of immigration). It closed with a second wave.....

Frederiksen's argument is that access to the "second wave" information would clarify the meaning of it, making an exhaustive search unnecessary. Since the results support the notion that all candidates were canvassed in both conditions, he concludes that this information played no role in resolution. This conclusion is, however, unwarranted. It is not possible to resolve it for the longer version without reference to the content of the second sentence; both period and America are possible antecedents. Thus, both top-down and bottom-up models predict that all candidates must be canvassed, even after the second clause has

been interpreted. They differ only in when the resolution process begins, but not with respect to how long it takes or when it ends.

5.2.4 Search processes

Springston (1975), noting that only reflexive and reciprocal pronouns are subject to grammatically-determined resolution (i.e., the Clausemate Constraint), hypothesized that non-clausemates need not be canvassed as candidate antecedents during resolution, since the search can self-terminate as soon as the clausemate is found. Conversely, he expected the search process for simple pronouns to be exhaustive, that is, to include all the NP's within the search domain, simply because there are many different positions in which the antecedent for a simple pronoun may occur.

Since Springston does not cite sources for the exhaustive vs. self-terminating distinction, let us first consider the character-classification studies which originally motivated the use of this terminology. Sternberg (1966) presented alphanumeric characters to subjects who were to decide as rapidly as possible whether each was a member of a previously-memorized set of up to six letters or numbers. Since response times for both positive and negative decisions increased linearly with the size of the memorized set, Sternberg inferred that all members of the reference set had been serially scanned. The rationale for this is simply that negative decisions can only be made when all candidates have been considered. If positive identifications take the same amount of time, they must also result from an exhaustive search. Theios (1973) characterizes this as a high-

speed scan, stating that

> in such situations it is convenient to assume that "copies of the memory representations of the recurring stimuli and responses may be held and organized into serially-scanned, short-term stores or memory buffers (p. 44)

Thus, the domain of serial-scanning searches is assumed to be working memory. If other items must be scanned, they must first be brought into active memory.

Springston's methodology and conclusions are founded on two basic assumptions. First, the use of a self-terminating search (as with reflexives) is entirely dependent on resolution being determinate; if an antecedent may occur in many different positions, a self-terminating search strategy would result in many errors, since the first possible antecedent would always be accepted. Secondly, when a candidate antecedent is canvassed, the speed with which it may be rejected is a function of the number of features on which it differs from the pronoun.

In a combined reading time - verification time task, Springston found faster comprehension for (268) than for (269):

268. John left the party because Peter hit him.

269. John left the party because Carol hit him.

This suggests that the search is exhaustive, since the Peter/Carol difference would not arise unless both items had been involved. The lack of such an effect for (270):

270. John left the party because Peter praised himself.

271. Carol left the party because Peter praised himself.

further suggests that for reflexives, the search ends at the clause-mate of the pronoun.

The fact that Springston collapsed reading times and reaction times to a verification task such as (272):

272. John was hit.

for each item poses problems of interpretation. As with Hirst and Brill's (1980) antecedent-naming task, the verification task forces subjects to consider all NP's in recent memory in order to decide whether or not they meet the criteria for acceptance. In the case of a simple pronoun, the fact that there are numerous exceptions to the so-called Converse Clausemate Constraint would, according to Springston's model, require that the verification search include all of the NP's in the sentence. In the case of a reflexive, however, a verification search may be terminated when a match involves a reflexive, provided that the processor has access to the surface code as well as the meaning code. In Springston's task, this is a reasonable assumption, since the verification sentence immediately followed the test sentence. Thus it is possible to explain the gender effect entirely in terms of the verification portion of the task.

On the other hand, with a top-down processing model, one would not expect to find a gender effect during on-line processing. Given the contextual constraints that develop across a sentence such as (268), an optimal processor would be able to predict during the processing of Peter (or Carol in 269) that if a personal pronoun appears in the clause, it must have as antecedent the current topic. Interactive processing would assign the correct antecedent without an exhaustive search of other nearby NP's. Thus, if one could separate reading times from verification times, one would expect to find the gender

effect only for the latter task.

5.2.5 Complexity and dominance

Springston also asked whether the gender effect might disappear when the pronoun precedes and commands the non-antecedent, since the syntactic cue alone could serve to eliminate the NP in the lower sentence as a possible antecedent:

273. He said that John/Carol was sick.

Violation of the exhaustive search assumption for simple pronouns might not be surprising in this case, since the processor could be sensitive to the futility of searching forward into a lower clause, just as it does not go beyond the leftmost clause boundary for reflexives. Even if we were to accept the validity of the combined reading and verification measure, however, Springston's methods for investigating this question are not adequate. Rather than simply test for a gender effect in items such as (273), he compared reading times for a set of sentences such as:

274. Bill claims that he was living in London when Joan was a movie star.

275. Bill claims that when he was living in London, Joan was a movie star.

276. Bill claims that Joan was a movie star when he was living in London.

277. Bill claims that when Joan was a movie star, he was living in London.

No interesting differences were found. Presumably, he was working under the assumption that for 274 the processor would have available on-line the information that the when clause could not contain an

antecedent for he, and that the search would therefore not go forward; in the other cases, the syntax does not block coreference, and the search would be exhaustive in both directions. The lack of differences is anomalous; both Schlesinger (1968) and Tanenhaus and Seidenberg (1981) were able to find such effects. However, the fact that each sentence contains the same number of potential antecedents explains the lack of difference, if we can attribute the effects only to the verification task. Furthermore, (274) is similar in structure to the "McIntosh" sentences of Chapter III; subjects are not sensitive to this putative blocking. These items do not test Springston's hypotheses, since in Experiment 1 it was found that similar items (8 and 28) were accepted by 40% of the speakers. Finally, differences in the semantic relations of (274) - (277) and (275) - (276) further complicate matters.

Springston's verification task is also equivalent to a forced-choice procedure, since in order to verify asserted coreference, one must consider all of the alternatives from the prior discourse. Access to these alternative NP's will be affected by how they are currently stored in memory, and therefore could be of interest for models of question answering, but since Springston's data also contains reading time information, it cannot be used to make inferences about either source of variation independently.

Springston's choice of methods is also surprising in view of the fact that he was a student of H. Clark, the originator of the reaction time to comprehension technique. Clark himself (Clark and Sengul, 1979) and others (e.g., Kaplan, 1981) have used this technique successfully to measure anaphora resolution times. It would therefore seem

possible to separate reading times from verification times simply by using the reaction time to comprehension instructions for the target sentence, and the usual verification instructions for the verification task. This is the technique employed in Chapter VII below.

Taken together with the literature on lexical and inter-sentential resolution biases, Springston's oft-cited work suggests that there now exists little reliable data concerning on-line search processes for this type of material.

5.2.6 Search domains

Two studies, Clark and Sengul (1979) and Kaplan (1981), have attempted to determine whether antecedent searches are speeded when the referent falls within some "privileged" search domain, such as the previous clause or sentence. Clark and Sengul phrase this in terms of a "discontinuity model" based on working memory effects:

> The premise of this model is that the...entities mentioned in the current sentence and one sentence back (have) a privileged place in working memory and so (are) readily available for examination, for search and identification...Referents will be identified readily only if they are mentioned in the current sentence or one sentence back. (p. 36)

Recall that both Frederiksen (1981) and Just and Carpenter (1978) provide strong support for the notion that the amount of intervening text (and in particular, the number of potential antecedents) between a pronoun and its antecedent affects resolution times. However, such a claim differs from the hypothesis that there exist syntactically-defined domains within which the search must be carried out.

Clark and Sengul's experiments test the hypothesis that reaction times to comprehension for pronoun- and definite-NP anaphora will

be faster when the antecedent appears within the privileged domain. Two points bear mentioning here. First, the notion that a prior clause or sentence has privileged status in memory is not an idea which is original to Clark and Sengul. Sachs (1967, 1974), Jarvella (1971, 1979) and Caplan (1972) are but three of the many researchers who have studied the role of syntactic boundaries in short-term memory for sentences. Secondly, Clark and Sengul's technique is not susceptible to the kinds of criticisms I have raised against the work of Springston, Ehrlich, and Hirst and Brill. Not only did they obtain an independent measure of processing time for the pronoun-containing sentence, but the search involved was contrived so as to always be based entirely on pragmatic factors occurring late in the sentence. In such materials the exhaustive search model is correct, and equivalent to the kind of search that goes on for a verification task.

Their initial hypothesis that the previous sentence constitutes the privileged domain was tested in simple descriptive paragraphs which allowed reordering of the sentences without affecting discourse cohesion. For both pronouns and definite NP's, comprehension times (and hence resolution times) were fastest when the antecedent appeared one sentence back, and equally slow (but only about 350 msec. slower) two and three sentences back:

278. A broadloom rug in rose and purple colours covered the floor. Dim light from a small brass lamp cast shadows on the walls. In one corner of the room was an upholstered chair. It appeared to be an antique.

These results support the general notion of a privileged search domain, but since each sentence in the search domain consisted of a single clause, the relative effects of sentence and clause boundaries could

not be tested. Materials were therefore designed in which the preceding clause did or did not end a sentence. In the following example, an NP two clauses back is either one or two sentences back:

279. ...next to a tree. The smell of coffee...The tree...

280. ... next to a tree, while the smell of coffee...The tree...

No effect was found: resolution was no faster when the second clause back was in the nearest sentence. Thus only the clause preceding the anaphor defines the privileged domain.

Kaplan (1981) contends that Clark and Sengul's materials did not contain adequate controls for the number of NP's in the context sentences. Under the exhaustive search assumption, resolution times should depend on the number of NP's canvassed, but as we see in (278), a large number of NP's preceded the target (e.g., rug, colours, floor), and this number varied from item to item. Furthermore, changes in the sentence order and placement of the antecedent meant that all of these NP's would precede both the antecedent and the anaphor in the clause one sentence back, but that a variable number would precede and follow the antecedent in the other conditions. Thus, if the number of NP's in the immediately preceding clause strained the limited capacity of working memory, searches of prior clauses would, on average, lie outside the privileged domain, simply because there exists an independent clause-bounded domain. Kaplan therefore used Springston's gender effect, as well as variations in number (singular and plural) and humanness (human and neuter) to determine whether NP's one clause back, two clauses back, or two sentences back are accessed in the same way. Consider these examples:

281. The waitresses laughed at the comedian.

282. The waitresses started laughing when the comedian began joking.

283. The waitresses started laughing. The comedian had begun joking.

284. ...He sounded like Woody Allen.

The pragmatically-controlled antecedent was always the last-mentioned NP, and the target always began with the pronoun. With the number of NP's in the search domain controlled in this way, target reading times were found to be identical across conditions; thus, the two preceding clauses both fell within the privileged domain, regardless of whether one of the clause boundaries was also a sentence boundary. Furthermore, feature manipulations on the NP in the first sentence affected resolution times additively.

Kaplan also inserted extra lexical content, either nominal or non-nominal, into test sentences such as:

285: Some hikers (who were curious) / (and a guide) walked through a forest. It was full of strange noises.

Here, reading times increased only with the addition of nominals to the earlier portion of the search domain, which suggests that the search is selective at least in that only NP's are canvassed.

These results cast serious doubt on the notion that the preceding clause constitutes a privileged domain for antecedent searches. Although limits for the domain were not identified, it seems that they should be based not on syntactic units such as the clause or sentence, but on the number of NP's active in working memory during an antecedent.

A number of recent studies of the reading process have exploited

the phenomenon of pronominal anaphora resolution to test hypotheses about the amount of processing carried out during a single eye fixation. These studies have important implications for the nature of antecedent searches.

The notion that antecedent searches are initiated immediately upon perception of a pronoun seems to be supported by Carpenter and Just's (1977) finding that 50% of pronouns initiated regressions back to the antecedent, either immediately after the fixation during which the pronoun was read, or at the end of the current line. Although Carpenter and Just concluded that pronouns are resolved either immediately or near clause boundaries, Ehrlich and Rayner (1983) point out that by asking subjects to determine whether each line was consistent with the preceding discourse, Carpenter and Just encouraged them to make regressions which might not otherwise occur. In fact, Ehrlich (1983) and Ehrlich and Rayner (1983) report regression rates of only 2 - 10% for pronouns under instructions to read for meaning.

Ehrlich and Rayner (1983) used the finding that the distance between a pronoun and its antecedent results in slower resolution (e.g., Clark and Sengul, 1979; Frederiksen, 1981; Just and Carpenter, 1978; Kaplan, 1981) to test the hypothesis that anaphora resolution is accomplished during the fixation on which the pronoun is encoded. Using materials which allow immediate resolution (given a top-down model):

286. ...Mark was mailing a copy of the paper to Susan (...) he/she was very...

they varied the distance from no intervening words (e.g. ... to Susan. She...) through an Intermediate condition (as in 286) to the Far condition, with several lines of text between the pronoun and its

antecedents. Pronouns themselves are seldom fixated, the authors first developed an algorithm for deciding which fixation included the encoding of the pronoun (p. 81), and measured its duration, as well as that of the immediately preceding and two later fixations. In all three conditions, the fixation which included pronoun encoding was longer than the preceding duration. This implies that the pronoun initiated heavier processing, most likely the antecedent search, as soon as it was encountered. However, fixation durations did not differ across the distance conditions, despite the fact that the Far condition is known to require more time overall. Instead, for the Near and Intermediate conditions, the increase in fixation duration was abolished on the next fixation, which suggests that resolution was complete by that point. In the Far condition, however, the slowed processing lasted through subsequent fixations as well, which indicates that anaphora resolution was continuing in parallel with later reading processes, causing a general processing slowdown.

While Ehrlich and Rayner are most concerned with the finding that full processing of a single word can extend over more than one fixation, let us consider the implications of this finding for anaphoric search processes in general. The fact that Far antecedents took longer to assign can be explained in terms of an exhaustive serial search in which additional intervening potential antecedents require more time to evaluate, as demonstrated by Kaplan's comparison of added NP's vs. other lexical material. Ehrlich and Rayner (1977, 1978) contrast cognitive lag theories of reading, which hold that there is a temporal lag between fixation and full semantic processing, with

process monitoring theories, which claim that processing occurs only during the current fixation. Just and Carpenter's (1980) reading model is an extreme instance of the process monitoring position in that it assumes that readers not only attempt to process each word as fully as possible (including lexical access, disambiguation, referent instantiation, and discourse function), but also that they do not move on to the next fixation until all such processing is complete. These results support cognitive lag models.

One advantage of eye movement data over reading time data is that in the former case one can determine rather directly where during comprehension of the sentence anaphora resolution begins, while in the latter case one has only a global measure of increased processing. This is not to say, however, that reading times cannot tell us anything about the locus of onset of resolution. In circumstances where the alternative to immediate resolution is a delay until the next clause boundary is reached, the on-line and clausal processing models make different predictions about global as well as local processing (see the discussion of Caramazza et al., 1977, in 5.2.2 above).

5.3 Reading times, verification times, and anaphora resolution

One of the most surprising implications of the exhaustive search model is that the human processor does not exploit incoming cues to resolution, but blindly examines each possible NP for gender and number features before moving "up" to take semantic and syntactic information into account. The work of Marslen-Wilson and Tyler (e.g., Tyler and Marslen-Wilson, 1980) strongly suggests that such information is

exploited during other types of on-line processing (see 5.4), and several studies reviewed here have pointed to similar effects in anaphora resolution (e.g., Caramazza et al, 1977; Tanenhaus and Seidenberg, 1981).

I have argued, however, that Springston's results may be attributed entirely to the verification phase of his combined reading time - verification time methodology. In this task, subjects must evaluate the truth of a statement which makes explicit the antecedent of a test pronoun, and it is possible that this amounts to a serial search of each of the nearby NP's. In this task, a gender or number mismatch should indeed facilitate the decision to eliminate from consideration an NP which is not the antecedent. Furthermore, verification should promote a careful re-evaluation of the meaning, and to whatever extent is possible, of the wording of the test sentence. Thus, resolution may be achieved very quickly for a sentence which occurs in a context that provides syntactic and semantic facilitation on-line, but later verification may involve processes which look very ~~much~~ like an exhaustive serial search.

In the following section I discuss these proposals in greater detail, as background to the specific resolution and verification models presented and tested in Chapter VI.

5.4 Anaphora resolution and the top-down bottom-up controversy

Several of the arguments developed in this chapter hinge on the question of whether contextual semantic and syntactic information can be exploited immediately to eliminate candidate antecedents. The exhaustive serial search model of Springston (1975), based on gender

facilitation, and expanded by Hirst and Brill (1980) to include scanning of later pragmatic information, represents a serial, bottom-up interpretation. The work of Frederiksen (1981) and Ehrlich and Rayner (1983) suggests that resolution occurs on-line, in parallel with other processes.

These conflicting positions reflect a more basic dichotomy between the two central theories of comprehension elaborated in the psycholinguistic literature. There are several ways in which they can be described: autonomous vs. interactive, serial vs. parallel, on-line vs. clause-by-clause, and top-down vs. bottom-up. The serial/autonomous/bottom-up "Clausal Processing" school (e.g., Caroll and Bever, 1976; Fodor, Bever and Garrett, 1974; Forster, 1974, 1978; Hurtig, 1978; Frazier and Fodor, 1978; and Cowart, 1982) maintains that the language processor is best modelled as a set of autonomous sub-components, each of which must complete its processing of the input from a lower level before the next highest level can make its contribution. This sequentially organized system reflects the autonomous organization of the subcomponents of a transformational grammar, and in its strongest form claims that during comprehension a listener first performs phonetic analysis to recognize phonemes, which then serve as input to the lexical (word recognition) component, which then identifies syntactic structure as input to semantic interpretation at the clause level, and finally takes into account contextual semantic cues to the intended meaning. The autonomy assumption states that facilitation effects are forbidden across components. For example, ambiguous clauses are first processed with all possible interpretations, but at

the end of this process, semantic information from context is used to disambiguate. However, context cannot be used during the processing of an ambiguous clause to bias interpretation on-line.

Evidence for the Clausal Processing hypothesis is weak, and currently consists of a series of critiques of the top-down school (e.g., Cowart, 1982) rather than a defence of its own methods. The well-known click studies and sentence-completion tasks (see Tyler, 1981, for a review) have been criticized for failing to tap comprehension on-line. In every case, the task involves post-processing responses, when what is needed is evidence that during processing there is no top-down interaction among the components of the processor.

In a variety of tasks involving both word recognition and clause-internal syntactic and semantic processing, Marslen-Wilson and Tyler have found evidence for on-line higher-level facilitation (e.g., Marslen-Wilson, 1973, 1975, 1976; Marslen-Wilson and Tyler, 1975, 1976; Tyler and Marslen-Wilson, 1977, 1980, 1982). For example, when shadowing recorded speech, subjects make errors which are always syntactically and semantically congruent with the structure of prior discourse, even if the error is made in mid-clause. Purposely distorted words on the tapes are often restored by subjects at delays as close as 250 msec. to a form which is consistent with the on-going discourse, a result which strongly supports the notion that the listener constructs a complete syntactic and semantic representation during the process of comprehension. In experiments using texts which are distorted semantically, syntactically, or in both ways, it is observed that lexical access, as measured by probe responses of various types, is

fastest for normal prose. Some of the studies relevant to these questions have involved anaphora resolution.

Tyler and Marslen-Wilson (1981) used several mispronunciation detection tasks to determine whether three types of anaphoric devices (superordinate, definite NP repetition, and pronoun) are equally effective. In this task, it is assumed that if an anaphoric device is successfully resolved, it will activate the meaning of the antecedent (see Chapter VII), and will have the same facilitating effect on recognition of subsequent related words as the original noun. With materials such as:

287. Context: Mother saw the postman coming from a distance.
Superordinate: The man brought a leffer from Uncle Charles.
Definite NP: The postman brought a leffer from Uncle Charles.
Pronoun: He brought a leffer from Uncle Charles.
Non-Anaphor: Mother brought a leffer from Uncle Charles.

subjects were required to press a timer button as soon as they perceived a mispronunciation (here, "leffer" for "letter"). For both adults and children of various ages, the three types of anaphor all facilitated mispronunciation detection as compared to the non-anaphor. This argues for immediate resolution, since only a full anaphoric interpretation would reactivate the meaning "postman", thereby facilitating lexical access to the related word letter. If the first segments of "leffer" begin the lexical access process for letter, then the detection of the mispronunciation should be rapid. However, in the non-anaphor condition, mother does not facilitate lexical access to letter, and responses represent a baseline level of priming for that word in this context.

Tyler and Marslen-Wilson (1982) examined the facilitation associated with repetitions, pronouns, and null anaphors:

288. Context: As Philip was walking back from the shop, he saw an old woman trip and fall flat on her face. She seemed unable to get up.

Repetition: Philip ran towards ... HIM/HER

Pronoun: He ran towards... HIM/HER

Null anaphor: Running towards...HIM/HER

The subjects heard the context sentences, and at the point indicated in the anaphoric sentence, saw one of two possible continuations, either consistent or inconsistent with the preceding sentences. They note that for the null anaphor, one must infer that the subject of running is Philip in order to determine, as with the other two cases, that the HIM continuation is inappropriate. However, they found no difference in response times across the three types, which seems to indicate that the inference took no measurable amount of time. Only the difference between appropriate and inappropriate probes made a difference in how long a subject took to name the pronoun (about 45 msec. longer for inappropriate probes).

These results are unexpected, since other studies of the inference process have found clear reading time differences for sentences which do and do not require inferencing. Clark and Haviland (1977) found that (291) took longer to process in the context of (289) than when it was preceded by (290):

289. Andrew was especially fond of beer.

290. We got some beer out of the fridge.

291. The beer was warm.

Tyler and Marslen-Wilson suggest that the finding of such an effect

is an artifact of the reading time procedure used in the study. However, they have no explanation of how such effects might arise if the inference process, the only apparent difference between the two contexts, adds no time.

Their conclusion, that reading time gives an incorrect view of on-line processing, depends on their ability to demonstrate that in the null anaphor condition, resolution was achieved on-line, just as it was in the two anaphor conditions. However, this is an erroneous assumption. The fact that the HIM continuation is inappropriate is not in itself evidence that resolution has been performed. Consider this continuation:

292. Running towards HIM / HER a passing doctor opened his black bag.

This demonstrates clearly that the HER bias is independent of anaphora resolution in the non-anaphoric case. HER continuations will be more likely after the verb running, since regardless of whether it is Philip or another person, the most likely person one would run toward in such an emergency is the fallen woman. This simple explanation of their results makes it unlikely that they have discovered a mysterious link between inference times and written presentation, and refutes their argument that global reading times cannot reflect real increases in processing time associated with on-line resolution.

Nonetheless, Tyler and Marslen-Wilson's otherwise admirable efforts to test the conflicting claims of top-down and bottom-up processing models have important implications for anaphora resolution. In particular, the general finding that all available sources of information are available on-line to the listener or reader allows

us to formulate some rather interesting re-interpretations of earlier models of the resolution process.

CHAPTER SIX

ON-LINE VS. POST-PROCESSING EFFECTS IN ANAPHORA RESOLUTION

6.1 Introduction

In Chapter 5, I claimed that the exhaustive search model of anaphora resolution may be inappropriate for cases in which contextual and syntactic cues become available on-line to facilitate the elimination of potential antecedents. Furthermore, support for the exhaustive search model for those materials which might show such effects has been derived from post-processing tasks such as verification and antecedent naming, each of which might in itself force a subject to conduct an exhaustive search not to resolve anaphora, but to verify his interpretation against the other NP's in the recent discourse, in accordance with the implicit demand to do so. Since on-line resolution and later verification are confounded in such tasks, neither provides evidence for the exhaustive search hypothesis, which could be attributed to either initial or later processing, or perhaps both.

In order to test the exhaustive search hypothesis, an experiment was designed in which anaphora resolution is facilitated by discourse topicalization, syntactic cues and morphological features. It is claimed that the gender and number features are often equiprobable

on-line, and that a reading time measure of anaphora resolution will not reveal such effects. It is further claimed that feature facilitation will be predictive of response latencies in a verification task, although not for the reasons usually assumed.

6.2 The Target reading model

In this section we shall examine the Target sentences used in this study from the perspective of a fully interactive top-down processor, as well as from the bottom-up perspective, in order to explain why the former model does not contain a feature facilitation component. To the extent that these predictions are borne out in the experiment, one can accept or reject the validity of the model.

The sentences in this study are of four related syntactic types, exhibiting all combinations of free, blocked, forward and backward pronominal anaphora. Each appears as the third element in a three-sentence discourse which establishes a discourse topic as a probable antecedent. The target sentence always contains one other NP which may or may not match the topic NP on the gender and number features. By comparing reading times for the Match and the Mismatch conditions, we can determine the role played by surface morphological features during resolution.

We shall consider each type in turn, based on the following Context sentences:

293. C1: The restaurant owner was shouting and screaming.

294. C2: SHE/HE was very drunk.

6.2.1 Free Forward Targets

The two forms of the free forward target sentences are represented by the following examples:

295. MATCH (SHE...) The daughter saw the bottle behind her.

296. MISMATCH (HE...) The daughters saw the bottle behind him.

The MATCH condition is used when C2 contains she, and the MISMATCH condition is used when it contains he. The NP daughter(s), therefore, may or may not match the pronoun (her or him) in gender and number.

Let us assume that the processor initially establishes restaurant owner as the topic of discourse. Sanford and Garrod (1981) have shown that a two-sentence context referring to the same character in both sentences is the minimal requirement for establishing a discourse topic which may be pronominalized preferentially. As the Match sentence is heard or read, a decision must be made as to whether daughter(s) constitutes a topic shift. This decision may be made at any time during the processing of the Target, up to and including the point at which the preposition is processed. By then, the processor has made a decision as to how to treat incoming feminine singular pronouns - either as resolved to the original topic, or, if a topic shift is perceived, as coreferential with the new topic. The important point is that the context prior to the pronoun is sufficient to make such a decision, which may stretch over several words. On the other hand, if the MISMATCH sentence is heard, the context is predictive of two possibilities, each of which may be decided upon in advance. At the preposition, either him or her may appear with equal likelihood, and since each results in a unique resolution, no processing difficulty is encountered.

Thus the assumption of full on-line interactive processing leads to the prediction of no reading time difference between (295) and (296).

The exhaustive search assumption makes a different prediction: when the pronoun is perceived, the processor canvasses both daughter(s) and restaurant owner (masculine or feminine). In the MATCH condition, each must be considered in the light of the rest of the sentence, which has not yet been given a fixed interpretation. Free forward anaphora is, at the sentence level, ambiguous, and according to the Clausal Processing Hypothesis, each interpretation in turn must be considered in relation to the prior context. This should lead to processing complexity as the merits of each interpretation are considered. For the MISMATCH condition, the pronoun is compared with each recent NP first in terms of surface features. The feature incompatibility between the pronoun and the mismatched candidate NP terminates that portion of the search quickly, and as a result, the reading time for the MISMATCH condition is faster.

6.2.2 Free Backward Targets

The Free Backward Targets are:

297. MATCH (SHE...) Behind her, the daughter saw the bottle.

298. MISMATCH (HE...) Behind him the daughters saw the bottle.

The top-down model predicts that resolution will occur on-line to the topic NP, and that after the prepositional phrase has been processed, the fact that a new subject NP is about to be heard or read leads to an immediate decision as to how to treat it. If no topic shift is expected, resolution is complete. The bottom-up model predicts that resolution will wait until later NP's have been processed. At the

end of the clause (after bottle), a backward serial search begins, and a mismatched NP is rejected more quickly than a matched NP.

6.2.3 Blocked Forward Targets

The targets are:

299. MATCH (SHE...) Behind the daughter she saw the bottle.

300. MISMATCH (HE...) Behind the daughters he saw the bottle.

The top-down prediction is simple: the processor exploits syntactic cues on-line, excluding daughter(s) from consideration as soon as it is realized that a context for a pronoun which may be coreferential with the topic NP has been created. This may happen at any point during processing of the adverbial. The exhaustive search assumption implies that the mismatched NP will be dismissed quickly, while the matched NP will require further input from the higher, syntactic level before it can be rejected. Thus, the MISMATCH sentence should be understood more quickly.

6.2.4 Blocked Backward Targets

The Blocked Backward Targets are:

301. MATCH (SHE...) She saw the bottle behind the daughter.

302. MISMATCH (HE...) He saw the bottle behind the daughters.

The top-down model predicts that the pronoun will be resolved immediately. As the prepositional phrase is perceived, the context becomes less predictive of a coreferential NP. According to Hirst and Brill (1980), however, the processor does not resolve until later information has been processed, even if it is in a position which disallows coreference. The feature mismatch should speed comprehension of (302) over (301).

6.5 The verification model

Since I have argued that tasks involving verification do not provide an appropriate test of the exhaustive search assumption for anaphora resolution, it would be useful to specify the model which explains how such tasks give rise to the observed facilitation effects. In this section, I will assume that morphological facilitation is a feature of verification tasks only, and derive predictions about the role of the surface cues of gender and number in verification. These will be compared with predictions based on the exhaustive search - feature match model.

In this study, verification sentences probe either the PP or the main clause, depending on where the pronoun is located. To simplify matters, let us consider only responses to sample items which are false, given topic resolution.

Direction of anaphora should have no effect, since it does not affect the truth value of the items. We may therefore further restrict discussion to one example each of Free and Blocked anaphora.

Consider first a Free case:

303. (THE RESTAURANT OWNER...SHE...)

304. Behind her, the daughter saw the bottle. (MATCH)

305. The bottle was behind the daughter. (FALSE)

306. (THE RESTAURANT OWNER...HE...)

307. Behind him, the daughters saw the bottle (MISMATCH)

[redacted]ottle was behind the daughters.(FALSE)

[redacted]lain why a mismatch should lead to faster rejection.

According to Kintsch's model of memory for text (Kintsch, 1974; Kintsch and Vipond, 1978), the target sentence is stored as a set of propositions such as SEE (DAUGHTER, BOTTLE) (i); and BEHIND (BOTTLE, RESTAURANT OWNER fem.) (ii). In this representation, the pronoun has already been interpreted as an argument of BEHIND. The verification sentence is also assigned a propositional representation: BEHIND (BOTTLE, DAUGHTER) (iii).

The verification task requires that the subject retrieve Target propositions in order to determine whether they match the verification proposition. The first step is to locate propositions with the same predicate as the verification sentence, namely, BEHIND(X). Thus, (ii) will be rejected and (iii) will be examined. Next, each argument must be verified. In this case, an argument mismatch is identified: DAUGHTER and RESTAURANT OWNER (feminine) are different concepts. Therefore, the correct decision is to reject the test sentence. For the second example, the same propositions are compared, with the exception that DAUGHTERS and RESTAURANT OWNER (masculine) further differ in gender and number. One explanation of facilitation, then, is that the gender and number mismatch between DAUGHTERS and RESTAURANT OWNER (masculine) speeds rejection of that proposition, compared with the concept mismatch between DAUGHTER and RESTAURANT OWNER (feminine).

Straightforward as this may seem, it is probably incorrect. First, it seems unlikely that the time required to notice a lack of correspondence between two very different concepts will be significantly affected by their morphology, although this is open to empirical investigation. Secondly, and more convincingly, subjects in these experiments hear

or read the verification sentences immediately after the Targets, and thus still have access to surface codes. In a verification task, which challenges one's interpretation of a sentence, one should attempt to retrieve as much of the original as possible for reprocessing. For free anaphora, this means that a subject will have access to a surface form which is in fact ambiguous, and will be able to extract the alternative representation, BEHIND(BOTTLE,DAUGHTER). Since this matches the verification, there is one true reading and one false one, and the two alternatives - already dealt with during the comprehension phase - must be compared. Considerations such as the likelihood that the alternative antecedent provoked a topic shift must be weighed in relation to the available surface form. This reprocessing will take time, and verification of a MATCH sentence will therefore take longer than verification of a MISMATCH sentence.

The claim that the gender effect will arise from a resurrection of ambiguity from surface representations in memory leads to another prediction. Blocked anaphora does not allow an alternative interpretation. Thus, the only difference between the propositional matching procedure for MATCH and MISMATCH blocked targets is at the level of argument matching. Referring to our examples, we see that for the following blocked case:

309. THE RESTAURANT OWNER...SHE...

310. She saw the bottle behind the daughter. (MATCH)

311. The daughter saw the bottle (FALSE).

312. THE RESTAURANT OWNER...HE

313. He saw the bottle behind the daughters. (MISMATCH)

314. The daughters saw the bottle. (FALSE)

the only comparison is SEE(DAUGHTER,BOTTLE) vs. SEE(RESTAURANT OWNER (fem.)BOTTLE) for the MATCH condition, and SEE(DAUGHTERS,BOTTLE) vs. SEE(RESTAURANT OWNER (masc.) BOTTLE) for the MISMATCH condition. No syntactic ambiguities complicate the issue, and the only difference between the MATCH and the MISMATCH condition has to do with argument matching. I suggested above that verifying the identity of DAUGHTER against RESTAURANT OWNER (fem.) and of DAUGHTERS against RESTAURANT OWNER (masc.) is a matter for empirical investigation. If we find the predicted gender effect for Free verifications, we may compare the facilitation to that found in Blocked verifications. If the effect is large, we have evidence that the argument match is significantly affected by morphological features. If there is no effect in the blocked cases, we can assume that argument matches in verification are based on essential semantic features of the words.

EXPERIMENT 2

Method

Stimulus passages and design. A master list of 64 four-sentence test passages was written. Each passage consisted of two context sentences, a target (pronoun-containing) sentence, and a verification sentence. The context sentences served to foreground one character as the topic of discourse. Initial mention in the first context sentence used a gender-neutral term (e.g., singer), while in the second sentence gender was specified by use of a personal pronoun. Topics were

male in half the items and female in the other half.

The target sentences were of eight types, representing a complete crossing of three factors: Blocking (2 levels), Direction (2 levels), and Gender/number Match (2 levels). Examples are given in 6.2 and 6.3 above. Eight passages were written for each combination of conditions.

Half of the verification sentences were true, specifying coreference with the topic NP of the Context sentences, and half were false, specifying coreference with the other NP in the same sentence. For each passage, the verification question therefore differed between the Free and Blocked conditions. Furthermore, for two of the eight sets of conditions, the Free Match conditions, the truth value of the answer depended on the subject's interpretation of the ambiguous coreference of the Target. This means that error rates cannot be computed for those items. The Master list is reproduced in Appendix D.

From the master list, seven other lists were derived using a Latin square layout, such that each item on the master list appeared on each other list under a different set of conditions.

Sixty-four distractor passages were also used. These varied in length and content of the sentences, target type, and verification type. For many of the targets containing pronouns, the verification task tested a different aspect of the sentence. Fifteen practice passages, similar to the distractor passages, were also prepared. The practice and distractor passages were identical for all subjects.

The eight experimental conditions were crossed with eight groups of subjects (four per group) and eight sets of sentences (eight per condition), in a Latin square. This yielded 256 observations per

condition. Every subject therefore saw eight items per condition, and received each passage under one of the conditions.

Procedure. Subjects were assigned at random to the groups. They were tested individually in a soundprotected room with subdued lighting, seated in front of an Apple II+ computer, approximately 25 cm. from an Electrohome ESM 914 black and white monitor. A Cyborg ISAAC 91A data acquisition system controlled the timing of the responses.

Stimulus presentation was controlled by SPART (Stimulus Presentation And Reaction Time), a programme written in Apple Basic and Labsoft (Zegarra and Smyth, 1984). A listing is provided in Appendix E.

In a SPART-controlled experiment, each session begins with a ***RELAX*** signal centred in the upper half of the screen. This signal also appears between passages. Subjects bring sentences onto the screen by pressing the space bar, marked "Next Sentence", with the thumb. This starts the timer. With each successive depression of the space bar, the timer stops, the response and response time are recorded, and the timer is cleared; this occurs during the approximately 10 msec. required to clear the screen and print the next sentence. Timing for the next sentence begins when it has been written on the screen.

When the second Context sentence comes up, the subject sees a line of dots indicating that the Target sentence is about to appear. When the space bar is depressed this time, the screen goes blank for approximately 400 msec.. After reading the target, the subject presses the space bar again to view the Verification sentence, which is printed in inverse print - dark letters against an illuminated

background. Responses are made by pressing the V and N keys, labelled "true" and "false" respectively, with the index fingers. (A cardboard template covers the rest of the keyboard). The end of a passage is indicated by the ***RELAX*** signal. The programme presents the passages in a different random order for each subject.

Subjects were first read the standard instructions reproduced in Appendix F, then led carefully through the first four practice passages. The instructions emphasize that they should read each sentence at a normal rate, and should go on to the next sentence only when the preceding one has been fully comprehended. They were told that the Target sentence was being timed, and that they were to respond at the moment they felt they had understood it, as a reflection of their normal reading rate. They were also informed that there were three prizes of $25, $15 and $5 for the best performance on the true/false items.

During the practice session, subjects were instructed not to look at their fingers when responding. If necessary, practice trials were repeated until the subject felt comfortable with the task.

The experimenter remained with the subject throughout the practice session, but left the room during the test session. Subjects were told that there would be a break after half the trials, and that they would then be questioned about their accuracy and ability to follow the reading time instructions. At the break, each subject was asked if s/he could do better for the second half. They were encouraged to make use of the ***RELAX*** signal after error trials and when they felt the effects of fatigue.

Total running time was approximately 45 minutes per subject.

Results

Pre-analysis. SPART replaced latencies exceeding 10 seconds with that value. This accounted for 1.32% of the data. Means, standard deviations, skewness, kurtosis, maxima and minima were computed for each subject and for each item. The data were highly variable, especially in terms of skewness. Reaction time data are usually positively skewed, since there is a lower limit, but no upper limit on response times. Long latencies can be attributed either to difficulty or to error; faster latencies are more likely to represent the phenomenon of interest, while long latencies are more likely to represent experimental error.

As an alternative to the data trimming procedures suggested by Winer (1971) and criticized by Shoben (1982), median scores were computed and used as input to the ANOVA's. Error latencies for the six verification conditions which lend themselves to such analysis were replaced by the procedure suggested by Winer (1971). Error rates may be found in Table 13.

Target reading times. Analysis of variance was performed on medians per condition for each subject and on medians per condition for each test paragraph. The overall F was significant ($F(1,28) = 496.3$, $p < .001$; $F(1,7) = 328.77$, $p < .001$). Notably, the Match factor was not significant ($F(1,28) < 1$; $F(1,7) < 1$), nor were any of its interactions. Only the Blocking by Direction interaction attained significance

CONDITION	READING TIME	GENDER EFFECT
FFM	2255	+81 n.s.
FFMM	2174	
FBM	2384	-114 n.s.
FBMM	2498	
BFM	2322	-18 n.s.
BFMM	2340	
BBM	2117	+21 n.s.
BBMM	2096	

Table 11. Target reading times per condition in milliseconds, Experiment 2. Based on medians. Key: first letter: F = free, B = blocked; second letter: F = forward, B = backward; third letter(s): M = match; MM = mismatch.

	FORWARD	BACKWARD	DIRECTION EFFECT
FREE	2215	2441	-226 **
BLOCKED	2331	2107	+224 **
Blocking effect	-116 *	+334 **	

Table 12. Blocking by Direction interaction.

($F(1,28) = 24.2$, $p < .001$; $F(1,7) = 13.9$, $p < .01$). Table 12 shows the results of a Newman-Keuls test. Forward anaphora sentences were read more quickly for the free than the blocked cases, while backward anaphora was easier for the blocked cases ($p < .01$).

Verification times. Verification times and error rates for each condition are shown in Table 13. Note that error rates cannot be computed for the two free match conditions, since either response may

CONDITION	VERIFICATION TIME	ERROR RATE
FFM	2356	**
FFMM	2037	5.1%
FBM	2600	**
FBMM	2308	12.9%
BFM	1834	5.5%
BFMM	1790	11.0%
BBM	1928	5.1%
BBMM	1822	4.7%

Table 13. Verification times and error rates per condition, Experiment 2, based on medians. ** error rates cannot be determined (see text). For codes, see Table 11.

	FREE	BLOCKED
MATCH	2478	1881
MISMATCH	2173	1806
Match effect	305**	75 n.s.

Table 14. Match by Blocking interaction, Experiment 2, Verification, based on medians. Times in milliseconds. Free vs. blocked differences are artifactual (see text).

be considered correct.

The overall F was significant ($F(1,7) = 641.9$, $p < .001$; $F(1,28) = 443.8$, $p < .0001$). The main effect of gender/number Matching was also significant ($F(1,7) = 36.35$, $p = .0005$; $F(1,28) = 33.7$, $p < .001$). An

artifactual main effect of blocking also reached significance ($F(1,7) = 11.87$, $p = .011$; $F(1,28) = 11.31$, $p = .0022$). Verification sentences were always shorter for the blocked versions of the Targets, hence the faster verification times.

There was, however, a real interaction between Match and Blocking ($F(1,7) = 11.05$, $p < .0127$; $F(1,28) = 12.86$, $p = .0013$). This can be seen in Table 14, which presents the results of a Newman-Keuls test. The Match factor was significant only for the free targets ($p < .01$).

Discussion

Analysis of gender and number facilitation revealed no effects whatsoever in any of the conditions in the Target reading time data. However, this cannot be attributed to lack of sensitivity of the design, since clear effects were discovered for both the Direction of anaphora and for the Free - Blocked distinction.

The significant reading time differences can be taken as a measure of the naturalness or ease of processing. Keeping in mind that all sentences involved a main clause and an adverbial PP, we can say that free anaphora is less natural when backward than when forward, even when the backward cases are in fact resolved to a prior topic of discourse. Thus, sentence such as (314) are read more quickly than sentences such as (315):

314. (Topic: FRED) John saw a snake in front of him.

315. (Topic: FRED) In front of him, John saw a snake.

This extends the finding of more difficult processing for "indeterminate words" (Schlesinger, 1968) in an interesting way. Schlesinger found

that such sentences took longer to read when in isolation, and assumed that this could be attributed to the fact that the pronoun had to be held unresolved in working memory until an antecedent was found. However, the same is true when the pronoun is resolved immediately to a discourse topic: the processing difference is a general characteristic of this sentence type. The implication for Schlesinger's work, as well as for Springston's attempt to examine the relative contributions of gender and precede-and-command to reading times, is that factors other than determinacy of the pronoun can affect reading times. One possibility is that the free backward cases are less frequent than the free forward. Thus a comparison across types does not necessarily reflect pronoun processing alone.

For blocked anaphora, the reverse is true: blocked forward cases took longer to process than blocked backward:

316. (FRED) He saw a snake in front of John.

317. (FRED) In front of John he saw a snake.

In this case the comparison across directions holds the direction of resolution (to the topic) constant.

Comparisons across free and blocked cases are perhaps less important, since they have different meanings. Free anaphora was easier than blocked in these contexts for the Forward cases (314 vs. 317) but Blocked was easier for the backward cases (315 vs. 316).

Most importantly, these results fully support the contention that anaphora resolution during normal reading in discourse is not a backward serial search process analogous to Sternberg's (1969) memory scan model. Instead, the resolution model based on Tyler and

Marslen-Wilson's interactive parallel approach is supported by these results. The lack of gender/number facilitation is a consequence of the interaction of syntactic, semantic and discourse principles during processing, which renders the morphological cues redundant.

The overall approach grew out of Marslen-Wilson's (1973) early observation that even at shadowing lags of only 250 msec. between input and output, subjects' errors showed definite evidence of on-line construction of a complete lexical, syntactic, and semantic representation of the incoming speech. Given that output processes must have accounted for the majority of the 250 msec., comprehension must be very fast and very complete indeed. Therefore, the complex assumptions concerning the interactions among subcomponents of the processing system are not incompatible with the observed speed and complexity of linguistic processes.

It should be noted that this model does not imply that anaphora resolution is <u>never</u> a backward serial search process. Under certain conditions, namely, when the only key to resolution is a pragmatic clue which follows the pronoun, on-line access to resolution information is absent, and the resolution process becomes a serial match routine. Kaplan's (1981) materials are a good example:

318. The waiter/waitress laughed at the comedian.

319. <u>He</u> sounded like Woody Allen.

Nonetheless, here the on-line interactive model does predict immediate resolution for the MISMATCH condition, but delayed resolution for the pragmatically resolved MATCH condition. In this case, gender cues are not redundant, but serve to facilitate immediate processing.

The verification results are important for two reasons. First, they show that gender/number facilitation is a real effect, but when discourse information makes it redundant on-line, it can still be observed in verification. This strongly suggests that the arguments against the use of verification and related measures (such as antecedent naming) to test hypotheses about comprehension are valid ones. In this instance, the effect has nothing to do with anaphora resolution.

The second important aspect of these results is that the verification model outlined above is also fully supported. Recall that the model is based on a combination of comparison processes for propositional representations, and access to residual surface codes which allow reprocessing and uncertainty concerning the initial understanding of reference relations in the target sentences. One of the predictions of this theory was that the ambiguity would not be possible for blocked sentences, and that the only match process involves argument comparisons. Since the MATCH vs. MISMATCH difference is non-significant, we may conclude that in an argument matching routine, the processor is affected more by the semantic encoding of the terms than by their surface morphological features.

In summary, the top-down, on-line interactive model of language comprehension has provided us with a new framework within which to interpret the anaphora resolution process. Previously inexplicable results, most particularly the insensitivity of antecedent searches to top-down effects, are seen to have been artifactual.

Nonetheless, the lack of contextual facilitation observed here

should be treated with caution. Although we have seen good evidence that anaphora resolution is not affected by the kinds of relatively slow search processes claimed by others, it could be the case that these effects do exist, but at much faster speeds than can be measured with button-press reaction times. Only the exploration of such phenomena with increasingly sensitive tasks will allow us to draw increasingly firm conclusions about the facilitation of on-line processing by contextual information.

CHAPTER SEVEN

POST-RESOLUTION PROCESSES: THE ACTIVATION OF ANTECEDENTS

7.1 Introduction: Priming and activation

Semantic priming may be defined as facilitation of the processing of a word by the processing of a previous, related word. Foss (1982) cites an early description of the phenomenon from James' Principles of Psychology:

> Each word is doubly awakened; once from without by the lips of the talker, but already before that from within by the promonitory processes irradiating from the previous words" (1890, Volume I, p. 450)

James' characterization shows that psychologists had been aware of this aspect of top-down processing long before the recent popularity of this way of thinking about language. After nearly a hundred years, priming has finally gained the status of a central issue in three related domains: word recognition, comprehension processes, and semantic memory processes.

In this chapter we shall be especially concerned with priming effects insofar as they enable us to determine the state of activation of an antecedent after anaphora resolution has been accomplished. Activation of a concept may be viewed as a temporary change of state of a permanent memory item to a condition in which it is more able to participate in cognitive processing. Active memory is often referred

to as "working memory" (Baddeley and Hitch, 1974; Baddeley, 1980).

The relation between activation and priming may be illustrated by reference to the lexical decision task, one of several priming tasks that have been used to explore the notion of relatedness among concepts in semantic memory. In lexical decision experiments, subjects are timed as they indicate their decision as to whether or not a given letter string is a word. If a test word is closely related in some way to an immediately preceding test word, positive decisions are faster than in a control condition where the preceding word is unrelated. This facilitation is explained as follows: when a test word is recognized, the associated concept in semantic memory is activated, and remains in this state for some time. Moreover, activation spreads to related concepts in the semantic network which constitutes long-term memory (e.g. Collins and Quillian, 1970; Collins and Loftus, 1975). Within the semantic network, related concepts share nodes (roughly, semantic features), and activation of the representation of one item also leaves all of the nodes it shares with other items in a state of activation. In addition, both Collins and Quillian (1975) and Anderson (1976) suggest that activation levels decrease in proportion to their distance from the initially activated concept; otherwise, activation of a single item would eventually result in activation of the entire semantic network, and you would throw up.

Activation and priming techniques were first used only to investigate the permanent structure of semantic knowledge representations. However, McKoon and Ratcliffe (1976, cited in Ratcliff and McKoon, 1978) show in a lexical decision task that newly learned asso-

ciations (e.g. city-grass) primed one another to the same extent as words with prior associations (e.g., green-grass). Ratcliff and McKoon (1978) extended this finding to learned sentences in an item recognition paradigm. Priming occurred when a test word was preceded by a word from the same sentence. In addition, if the test word was from the same proposition in that sentence, priming was stronger.

This latter result confirms a prediction derived from Kintsch's propositionally-based model of text comprehension and memory (e.g., Kintsch, 1974; Kintsch and Vipond, 1978). According to this model, the fundamental unit of text memory is the proposition rather than individual word concepts; thus, priming one concept in a proposition necessarily primes other concepts within the same proposition. Recently, Guindon and Kintsch (1984) have shown that similar effects are found for macropropositions, which are elements of a more global representation of the text that speakers construct during reading. They found larger priming effects for pairs of words taken from a summary statement at the end of a short paragraph, whether or not that statement had been overtly expressed in the text.

Foss (1982) pointed out that the earlier list experiments showed rapid decay times for priming effects, which continue only as long as the first word is being processed. Under the assumption that priming has a function in language comprehension, this is a puzzling result, for one would expect priming effects to act over fairly large stretches of text. Foss therefore used a phoneme monitoring task to investigate priming across word pairs in both coherent and scrambled sentences. He found that while priming

decreased sharply in the scrambled version as a function of the distance between the priming word and the critical word, no such decrement appeared for words in sentences. Foss interprets this in terms of a model of discourse comprehension in which core propositions remain active as long as the type of discourse does not change.

7.2 Priming and activation in anaphora resolution

Activation of antecedents has been examined in three studies based on Caplan's (1972) probe recognition task. In this paradigm, subjects read or listen to text, and at unpredictable times are presented with a probe word from prior text which they must judge "new" or "old". For a currently active word, recognition should be facilitated, but for deactivated words, extra retrieval time is needed to reactivate the concept, thereby making the decision possible.

Caplan suggested that clause boundary effects on the availability of information from working memory should correlate with activation. Jarvella (1971, 1979) describes several experiments which show better rote recall from an immediately preceding clause than from one clause further back, and Sachs (1967, 1974) has found that surface information is dismissed rapidly from memory for sentences within stories.

Caplan designed materials such as this example:

320. Whenever one telephones at night, rates are lower.
PROBE: night

321. Make your calls after six, because night rates are lower.
PROBE: night

He found significant clause boundary effects: probe latencies were slower when the probed word was two clauses back, even when the actual

words following the target were identical in the two condition.

Chang (1980) exploited this technique to determine whether a pronoun activates its antecedent. Consider these materials:

322. John and Mary went to the grocery store / and John bought a quart of milk. PROBE: JOHN

323. John and Mary went to the grocery store / and Mary bought a quart of milk. PROBE: JOHN

324. ...and he bought a quart of milk. PROBE: JOHN

325. ...and she bought a quart of milk. PROBE: JOHN

In each case, JOHN is the probe word for which subjects must provide an old/new judgment. In (322), probe latencies should be fast, since the repeated noun is now, by all accounts, in active memory. In (323), however, latencies should be slower, since JOHN appears only in the first clause. For (324) and (325), Chang asked whether probe latencies would parallel the full NP conditions. If so, this would demonstrate activation of the antecedent by the pronoun. By contrast, she should not activate John; recognition must be based on retrieval from the earlier clause just as in (323).

As it turns out, recognition was faster for the repeated noun condition than for the pronoun condition (i.e., 322 vs. 324), presumably because the repetition condition provides extra visual cues. Most importantly, both of these were faster than the item in which only she or Mary appeared in the most recent clause. Moreover, latencies were similar in each of the two latter cases. Thus, the pronoun he reactivated John in the most recent clause, and primed probe recognition.

Two papers pick up this theme by asking whether an anaphor acti-

vates other concepts along with the antecedents. McKoon and Ratcliff (1980) used Caplan's probe latency task to show that a definite NP anaphor (always a superordinate word) activates its antecedent, as in:

326. A burglar surveyed the garage set back from the street. (neutral continuation...)

327. The burglar / the criminal / a cat slipped away from the streetlamp.

PROBE: BURGLAR

As in Chang's study, priming effects were found for both the criminal and the burglar versions of the final sentence. Probe recognition for the unrelated cat version, however, was slower. This supports the conclusion that criminal served to reactivate burglar from the first sentence, thus priming burglar for probe recognition. When the probe was garage (which also appears in the first sentence), priming was again observed for the criminal condition. This serves to confirm that the probe effect for burglar was not caused solely by prior association with criminal, since this would not explain how garage could have primed the probe. The results also suggest that an anaphor activates both its antecedent and "companions" in the same proposition as the antecedent. This latter result is in fact predicted by the theory of text memory advanced by Kintsch and Vipond (1978).

Dell, McKoon and Ratcliff (1982) extended these results in a similar study, this time using word-by-word presentation, making it possible to determine when during processing an antecedent is activated, and for how long. It was shown that probe latencies are speeded for both the antecedent and words in the same proposition, as early as 250 msec. after the anaphoric item is read. No priming is found for

words in a different proposition taken from one of the neutral continuation sentences. Furthermore, activation faded more rapidly for the companion than for the antecedent. Specifically, companion priming effects disappeared at a probe point only 1250 msec. beyond the anaphor, while antecedent probe priming was still strongly in evidence.

7.3 Activation of antecedents across a clause boundary: Some residual questions

Chang (1980) observed different activation levels for an antecedent and a "control" item, which was in fact a companion, as proposition-mates are now termed. This result provides support for the notion that a pronoun activates its antecedent, but it does not tell us specifically how the difference in activation came about. Chang seems to have assumed that the control item represented baseline activation, but the work of McKoon and Ratcliff (1980) and of Dell, McKoon and Ratcliff (1982) suggests that companions should be activated along with the antecedent when a pronoun is resolved. On the other hand, companion activation is short-lived, disappearing by 1250 msec. post anaphor in their studies. Moreover, the propositional model upon which these predictions are based is intended to account for long-term representations of texts, and one could argue that there is an important difference between the work of Dell et al., who had the probe appear after three full sentences had intervened between it and the antecedent, and Chang, whose probe was timed to appear immediately at the offset of the test sentence. Many possibilities arise: perhaps for short-term representations, an antecedent is

selectively activated because the representation of the sentence has not passed from its surface form to its full propositional representation. This would imply that the sooner the probe appears after the clause boundary, the larger the difference one should find between antecedent and companion activation. On the other hand, it could be that despite their proximity to the antecedent, Chang's probes came far enough after the pronoun to fall outside the 1250 msec. region defined by Dell et al. This too would imply that the companion was fully deactivated, but only after having declined from a peak.

Fortunately, Chang published mean reading times per word as a function of serial position (p. 63); using these data we can estimate that for the examples cited above (322-325), the pronoun would have been read, on average, 1568 msec. before the appearance of the probe. This figure is more than 25% beyond the limits suggested by Dell et al., so that we may assume that the different activation levels did not represent partial activation. But we still do not know whether the companion could have been activated in the first place, given the recency of the test sentence.

To explore this issue, a partial replication of Chang's study was designed. There were two important differences. First, the second clause of the test sentence was reduced to an average of 3.9 words after the pronoun in order to allow the probe to appear within the domain of companion activation. This will allow us to detect any companion activation that would have been missed in Chang's study. Second, a true control condition was added to the antecedent probe and the companion probe conditions: a No Search condition, in which the

pronoun in the second clause is I. Using the first person pronoun will enable us to measure true baseline activation levels of the target NP's.

With this design, we may now ask whether the propositional model holds up in the case of memory representations which include surface codes, and whether the differences found (if any) between the antecedent and the companion represent an active - inactive contrast, or merely an active - partially active difference.

EXPERIMENT 3

Method

Materials and design. A master list of 36 two-clause sentences similar to those used by Chang (1980) was drawn up. In each case, the first clause had a conjoined subject made up of two names, one male and one female, equated for production frequency with reference to the norms of Battig and Montague (1969). Each name was used only once. In each sentence, the second clause began with and, but, so or when, divided equally across conditions. The subject of the second clause was either he, she or I.

A probe word, one of the proper names, was selected from each sentence. These were balanced for position (first or second member of the pair), and for gender. For the first 12 items, the probe was the antecedent of the pronoun. This is the Antecedent condition. For the next 12 items, the probe was the other proper name. This was the Companion condition. For the last 12 items, corresponding to the pronoun I, the probe was either NP, again balanced for position and

gender. There were an average of 3.9 words from the pronoun to the end of the sentence, and hence the beginning of the probe.

For each item, a verification item was also written. This tested randomly chosen aspects of the meaning of each item. There were equal numbers of true and false answers.

Fifty-four filler items were also used. Some contained pronouns, but probed a different word. Forty had false probes, testing words which were semantically related to a word in the sentence, or testing totally new words. One set of test sentences appears as Appendix G.

The three experimental conditions (Antecedent, Companion and No Search) were crossed with three groups of subjects (five per group) and three sets of passages (12 per set), in a Latin square design.

Subjects. The subjects were 15 adult English speakers, six male and nine female, ranging in age from 18 to 50 years (M = 28.4). Students were paid $4.00 for their participation; others participated voluntarily. One subject was replaced for not following instructions.

Procedure. Equipment and testing space were identical to those of Experiment 2. Stimulus presentation was controlled by an adapted version of SPART (Zegarra and Smyth, 1984). A listing is provided as Appendix H. Subjects saw a series of asterisks to begin. They pressed the space bar with the thumb to view the target sentence, and again to indicate the moment at which they had fully understood. This resulted in 400 msec. of blank screen, followed by a fixation point in the centre of the line on which the test sentence had appeared. Using the index fingers, left-handed subjects pressed the D and K keys, while right-handed subjects used the L and F keys to respond True and

False respectively. The keys were marked appropriately, and a cardboard template covered the rest of the keyboard.

During the 15 practice trials subjects were told to respond without looking at their fingers. If necessary, some practice trials were repeated until this condition could be met.

Results

Following Chang (1980) an analysis of variance was performed. The overall F was significant ($F(1,12) = 262.61$, $p < .001$). Means and per cent correct per condition are given in Table 15.

	MEAN	S.D.	% CORRECT
ANTECEDENT	740	47.4	95.5
COMPANION	770	47.3	98.8
NO SEARCH	834	50.18	95.5

Table 15. Mean probe latency in milliseconds, Experiment 3, with per cent correct and standard deviations

A priori comparisons showed no difference between the means for the Antecedent and the Companion conditions ($F(1,12) = 1.50$, $p = .244$). There was a highly significant different between probe latencies for Companion and No Search ($F(1,12) = 1969$, $p = .0008$).

Discussion

There were two clear results, First, the activation level of the Companion was much higher than that of the No Search probe condition. This demonstrates that the phenomenon of propositionally-

based activation holds even for relatively "fresh" representations which preserve a significant amount of surface information. Apparently, semantic integration to propositional form is essentially immediate. Secondly, the activation level of the Companion NP was not significantly lower than that of the Antecedent. However, the difference of 30 msec. was in the right direction. This suggests that as the distance from the pronoun to the probe decreases, the activation level of the Companion NP increases, presumably unrestricted. The fact that Chang found a slightly larger difference which attained significance indeed suggests that his longer distance between pronoun and probe resulted in a considerable decrement in the activation level of the Companion, as one might predict on the basis of Dell et al.'s results.

Another difference between this study and Chang's was the use of whole-sentence presentation. Chang used word-by-word presentation under the control of each subject, and timed each word. However, this experiment has demonstrated that the same kinds of results can be obtained with the simpler technique. It would appear that the visual presentation does not have an unnatural effect on probe recognition. The instructions to subjects to read each test sentence just once and to move on to the probe word as soon as the sentence had been understood appear to have been sufficient to ensure a normal range of fixations; furthermore, the finding of clear effects suggests that the distractor items fulfilled their purpose of breaking the development of artifactual strategies.

CHAPTER EIGHT

SUMMARY AND CONCLUSIONS

In this work I have examined a single linguistic phenomenon, English pronominal anaphora, from several different perspectives. First, the formal arguments for alternative grammatical representations were reviewed in order to establish a frame of reference for discussion of the role of cognitive strategies in judgments of the grammaticality of anaphoric and nonanaphoric sentences. An argument was developed concerning the nature of the judgment task itself: one can adopt either of two distinct criteria, namely, a well-formedness criterion and an interpretive criterion. In a judgment task, naive speakers listened to a sample of sentences which included 33 items which form the basis for linguistic arguments concerning the optimal grammatical representation of constraints on coreference. Over fifty per cent of the judgments were accepting of coreference on these putatively blocked sentences. Implications for the conduct of empirical research were discussed. Although two suggestions were made concerning data collection, that is, that production data be examined and accepted over judgment data for matters of well-formedness, and that judges be instructed to adopt production/well-formedness criteria, certain major difficulties remain. First, it is not clear how the finding that naive speakers do not access interpretive principles in a judgment task

should be understood. I have shown that linguists themselves have made the same confusion between well-formedness and interpretive constraints on anaphora, but the notion of interpretive principles is so closely allied to the notion of syntactic structure that it is not obvious that the inability to apply one of these exclusively in a judgment task should simply be brushed aside with a few methodological caveats. Secondly, the numerous cognitive strategies for interpreting coreference despite apparently well-motivated linguistic rules tend to blur the boundary between formal and functional approaches to linguistic description. It is particularly disturbing that for many of these strategies, no information processing model seems to have taken shape. Instead, we face a collection of vaguely psychologically motivated phenomena which are neither product nor process.

The second perspective on anaphora has been the view from cognition. Anaphora resolution has become a growth industry for those interested in modelling thought processes and integrating various levels of analysis into global theories of language comprehension. Anaphora is well-suited as a testing ground for such theories, as it cannot be given even the most cursory of characterizations without the inclusion of morphological, syntactic, discourse, pragmatic, and process details. I have tried to synthesize the experimental literature from the perspective of the top-down vs. bottom-up argument, contributing an experiment of my own whose goal it is to sort out the locus of the effects claimed to be representative of the resolution process in general. The finding that the evidence for the deployment of exhaustive serial searches is partly an artifact of inappropriate methodology

has helped clarify precisely what predictions one can make from a global theory to the description of a particular information processing task, in this case, on-line resolution and off-line verification. Preliminary models of the two processes were successful in accounting for the results of the resolution experiment.

Finally, I examined briefly the consequences of processing anaphora from the point of view of memory systems. A modest experiment demonstrated that propositional representation and its attendant effects on retrieval process remain an appropriate framework for even very short-term processing of language. Activation effects and the spread of activation to proposition-mates are as evident in an immediately preceding clause as they are at the level of macrostructure in much larger texts.

BIBLIOGRAPHY

Akmajian, A., and Jackendoff, R. Coreferentiality and stress. Linguistic Inquiry, 1970, 1, 124-126.

Anderson, J. Language, memory and thought. Hillsdale, N.J.: Lawrence Erlbaum, 1976.

Bach, E. Problominalization. Linguistic Inquiry, 1970, 1, 121-122.

Baddeley, A. The concept of working memory: A view of its current state and probable future development. Cognition, 1980.

Baddeley, A. and Hitch, G. Working memory. In G. Bower (Ed.) The psychology of learning and motivation, 1974, 8, 47-90.

Battig, W. and Montague, W. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monograph, 1969, 80 (3, Pt. 2).

Bever, T. The cognitive basis for linguistic structures. In J.R. Hayes (Ed.) Cognition and the development of language. New York: Wiley, 1970.

Bickerton, D. Some assertions about presuppositions about pronominalization. In Grossman, R., San, L. and Vance, T. (Eds.) Papers from the parasession on functionalism. Chicago: Chicago Linguistic Society, 1975.

Bily, M. Pronominalization rules (coreference rules) described in terms of Functional Sentence Perspective (FSP). In F. Karlsson (Ed.) Papers from the Third Scandinavian Conference of Linguists. Text Linguistics Research Group, Academy of Finland, 1977.

Bloomfield, L. Language. New York: Holt, 1933.

Bolinger, D. Judgments of grammaticality. Lingua, 1968, 21, 34-40.

Bolinger, D. Pronouns and repeated nouns. Indiana University Linguistics Club, 1977.

Bolinger, D. Pronouns in discourse. In T. Givon (Ed.) Syntax and Semantics 12. New York: Academic Press, 1979.

Bresnan, J. An argument against pronominalization. Linguistic Inquiry, 1970, 1, 122-123.

Bresnan, J. A realistic transformational grammar. In M. Halle, J. Bresnan and G. Miller (Eds.) Linguistic theory and psychological reality. Cambridge: M.I.T. Press, 1978.

Caplan, D. Clause boundaries and recognition latencies for words in sentences. Perception and psychophysics, 1972, 12, 73-76.

Caramazza, A., Grober, E., Garvey, C. and Yates, J. Comprehension of anaphoric pronouns. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 601-609.

Caramazza, A. and Gupta, S. The roles of topicalization, parallel function and verb semantics in the interpretation of pronouns. Linguistics, 1979, 3, 497-518.

Carden, G. Backwards anaphora in discourse context. Unpublished manuscript, Yale University, 1978.

Carden, G. and Dieterich, T. Introspection, observation, and experiment: An example where experiment pays off. Journal of the Philosophy of Science Association, 1980, 2.

Caroll, J. and Bever, T. Sentence comprehension: A study in the relation of knowledge to perception. In E. Carterette and M. Friedman (Eds.) The handbook of perception. Volume 5: Language and Speech. New York: Academic Press, 1976.

Carpenter, P. and Just, M. Comprehension as eyes see it. In M. Just and P. Carpenter (Eds.) Cognitive processes in comprehension. Hillsdale, N.J.: Erlbaum, 1977.

Carpenter, P. and Just, M. What your eyes do while your mind is reading. In K. Rayner (Ed.) Eye movements in reading: Perceptual and language processes. New York: Academic Press, 1983.

Carpenter, P. Inference processes during reading: Reflections from eye fixations. In J. W. Senders, D.F. Fisher, and R.A. Monty (Eds.) Eye movements and psychological processes II. Hillsdale, N.J.: Erlbaum, 1978.

Chafe, W. Givenness, contrastiveness, definiteness, subjects and topics. In C. Li (Ed.) Subject and topic. New York: Academic Press, 1976.

Chang, F. Active memory processes in visual sentence comprehension: Clause effects and pronominal reference. Memory and Cognition, 1980, 8, 58-64.

Chomsky, C. The acquisition of syntax in children from 5 to 10. Cambridge, Mass.: The M.I.T. Press, 1969.

Chomsky, N. Syntactic structures. The Hague: Mouton, 1957.

Chomsky, N. Aspects of the theory of syntax. Cambridge: The M.I.T. Press, 1965.

Chomsky, N. Deep structure, surface structure, and semantic interpretation. In D. Steinberg and L. Jakobovits (Eds.) Semantics. London: Cambridge University Press, 1971.

Chomsky, N. On binding. Linguistic Inquiry, 1980, 11, 1-46.

Chomsky, N. Lectures on government and binding. Dordrecht: Foris, 1981.

Clark, H. The language as fixed effects fallacy. JVLVB, 1973, 12, 335-359.

Clark, H. Inferences in comprehension. In D. Laberge and S. Samuels (Eds.) Basic processes in reading: Perception and comprehension. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Clark, H. and Haviland, S. Comprehension and the given-new contract. In R. Freedle (Ed.) Discourse production and comprehension. Norwood, N.J.: Ablex, 1974.

Clark, H. and Sengul, C. In search of referents for nouns and pronouns. Memory and Cognition, 1979, 7, 35-41.

Collins, A. and Loftus, E. A spreading-activation theory of semantic processing. Psychological Review, 1975, 82, 407-428.

Collins, A. and Quillian, M. Facilitating retrieval from semantic memory: The effect of repeating part of an inference. Acta Psychologica, 1970, 33, 304-314.

Collins, A. and Quillian, M. Experiments on semantic memory and language comprehension. In L. Gregg (Ed.) Cognition in learning and memory. New York: Wiley, 1972.

Cowart, W. Autonomy and interaction in the language processing system: A reply to Marslen-Wilson and Tyler. Cognition, 1982, 12, 109-117.

Culicover, P. A constraint on coreferentiality. Foundations of language, 1976, 14.

Delisle, G. Discourse and backward pronominalization. Bloomington: Indiana University Linguistics Club, 1973.

Dell, G., McKoon, G. and Ratcliff, R. The activation of antecedent information during the processing of anaphoric reference in reading. JVLVB, 1982, 22, 121-132.

Dinsmore, John. Pragmatics, formal theory, and the analysis of presupposition. Ph.D. thesis, University of California, Berkeley, 1979. Reprinted by the Indiana University Linguistics Club, 1981.

Dougherty, R. An interpretative theory of pronominal reference. Foundations of language, 1969, 5, 488-519.

Ehrlich, K. Comprehension of pronouns. Quarterly Journal of Experimental Psychology, 1980, 3, 247-255.

Ehrlich, K. Eye movements in pronoun assignment: A study of sentence integration. In K. Rayner (Ed.) Eye movements in reading: Perceptual and language processes. New York: Academic Press, 1983.

Ehrlich, K. and Rayner, K. Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. JVLVB, 1983, 23, 1-13.

Fodor, J., Bever, T. and Garrett, M. The psychology of language. New York: McGraw-Hill, 1974.

Forster, K. The role of semantic hypotheses in sentence processing. In Colloques Internationaux du CNRS No 206, Problèmes actuels en psycholinguistique. Paris, 1974.

Forster, K. Levels of processing and the structure of the language processor. In Cooper, W.E. and Walker, E. (Eds.) Sentence processing: Contemporary studies in psycholinguistics. Studies presented to Merrill Garrett. Hillsdale, N.J.: L.E.A., 1979.

Foss, D. A discourse on semantic priming. Cognitive Psychology, 1982, 14, 590-607.

Frazier, L. and Fodor, J.D. The sausage machine: A new two-stage model of the parser. Cognition, 1978, 6, 291-325.

Frederiksen, J. Understanding anaphora: Rules used by readers in assigning pronominal referents. Discourse processes, 1981, 4, 323-347.

Garvey, C. and Caramazza, A. Implicit causality in verbs. Linguistic Inquiry, 1974, 5, 459-464.

Garvey, C., Caramazza, A. and Yates, J. Factors influencing assignment of pronoun antecedents. Cognition, 1976, 3, 227-243.

Gleason, H. An introduction to descriptive linguistics. New York: Holt, Rinehart and Winston, 1955.

Grice, H.P. Logic and conversation. In P. Cole and J. Morgan (Eds.) Syntax and semantics, vol. 3: Speech Acts. New York: Seminar, 1975.

Grober, E., Beardsley, W. and Caramazza, A. Parallel function strategy in pronoun assignment. Cognition, 1978, 6, 117-133.

Guindon, R. and Kintsch, W. Priming macropropositions: Evidence for the primacy of macropropositions in the memory for text. JVLVB, 1984, 23, 508-518.

Halliday, M.A.K., and Hasan, R. Cohesion in English. London: Longmans, 1976.

Hankamer, J. and Sag, I. Deep and surface anaphora. Linguistic Inquiry, 1976, 7, 391-428.

Hill, A. Grammaticality. Word, 1961, 17, 1-10.

Hirst, W. and Brill, G. Contextual aspects of pronoun assignment. JVLVB, 1980, 19, 168-175.

Hockett, C. A course in modern linguistics. New York: The MacMillan Company, 1958.

Householder, F. On arguments from asterisks. Foundations of Language, 1973, 10, 365-376.

Hurtig, R. The validity of clausal processing strategies at the discourse level. Discourse Processes, 1978, 1, 195-202.

Jackendoff, R. An interpretive theory of pronouns and reflexives. Bloomington: Indiana University Linguistics Club, 1968.

Jackendoff, R. Semantics in generative grammar. Cambridge: The M.I.T. Press, 1972.

Jackendoff, R. An interpretive theory of pronouns and reflexives. PEGS Report no. 27. Washington, D.C.: Center for Applied Linguistics, 1973.

Jarvella, R. Syntactic processing of connected speech. JVLVB, 1971, 10, 409-416.

Jarvella, R. Immediate memory and discourse processing. In B. Bower (Ed.) The psychology of learning and motivation, 13. New York: Academic Press, 1979.

Jarvella, R. and Herman, S. Clause structure of sentences and speech processing. Perception and Psychophysics, 1972, 11, 381-384.

Jespersen, O. Language: Its nature, development and origin. New York: Holt, 1922.

Just, M. and Carpenter, P. A theory of reading: From eye fixations to comprehension. Psychological Review, 1980, 87.

Kaplan, B. Search processes in pronoun reference resolution. Unpublished doctoral dissertation, New York University, 1981.

Karttunen, L. Discourse referents. Mimeo, Indiana University Linguistics Club, 1969. Reprinted in McCawley (1976).

Katz, J. and Fodor, J. The structure of a semantic theory. Language, 1963, 39, 170-210.

Katz, J. and Postal, P. An integrated theory of linguistic descriptions. Cambridge: The M.I.T. Press, 1964.

Keenan, E. Anaphora and cross referencing systems. Paper presented at the Typology and Field Work Workshop, Oswego, New York, 1976.

Kimball, J. Seven principles of surface structure parsing in natural language. In Syntax and Semantics, 1973, vol. 1. Reprinted in Cognition, 2, 16-47.

Kintsch, W. The representation of meaning in memory. Hillsdale, N.J.: Erlbaum, 1974.

Kintsch, W. and Vipond, D. Reading comprehension and readability in educational practice and psychological theory. In G.L. Nilsson (Ed.) Memory: Processes and problems. Hillsdale, N.J.: Erlbaum, 1978.

Klima, E. Negation in English. In J. Fodor and J. Katz (Eds.) The structure of language. Englewood Cliffs, N.J.: Prentice-Hall, 1964.

Kuno, S. Functional sentence perspective: A case study from Japanese and English. Linguistic Inquiry, 1972a, 3, 333-378.

Kuno, S. Pronominalization, reflexivization and direct discourse. Linguistic Inquiry, 1972b, 3, 161-195.

Kuno, S. Three perspectives in the functional approach to syntax. In R. Grossman, L. San, and T. Vance. Papers from the para-session on functionalism. Chicago: Chicago Linguistics Society, 1975.

Kuroda, S. English relativization and certain related problems. Language, 1968, 44, 244-266.

Lakoff, G. Pronouns and reference. Bloomington: Indiana University Linguistics Club, 1968.

Langacker, R. Pronominalization and the chain of command. In Reibel and Schane, 1969.

Lasnik, H. Remarks on coreference. Linguistic Analysis, 1976, 2, 1-22.

Lees, R. and Klima, E. Rules for English pronominalization. Language, 1963, 39, 17-28.

Lesgold, A. Pronominalization: A device for unifying sentences in memory. JVLVB, 1972, 11, 316-323.

Lockman, A. and Klappholz, A. Toward a procedural model of contextual reference resolution. Discourse Processes, 1980, 3, 25-71.

Lust, B. Constraint on anaphora in early child language. Paper presented at the Second Annual Boston University Conference on

language development. October 1, 1977.

Lust, B., and Clifford, C. The 3D study: Effects of depth, distance and directionality on children's acquisition of anaphora. In Lust (B. (Ed.) Studies in first language acquisition of anaphora: defining the constraints.

Lust, B., Loveland, K. and Kornet, D. The development of anaphora: Syntactic and pragmatic constraints. Linguistic Analysis, 1980, 6(4), 359-390.

Lyons, J. Semantics. Cambridge: Cambridge University Press, 1977.

Maclay, H. and Sleator, M. Responses to language: judgments of grammaticalness. International Journal of American Linguistics, 1960, 26, 275.

Marslen-Wilson, W. Linguistic structure and speech shadowing at very short latencies. Nature, 1973a, 244, 522-523.

Marslen-Wilson, W. Speech shadowing and speech perception. Unpublished Ph.D. Thesis, Department of Psychology, M.I.T., 1973.

Marslen-Wilson, W. Sentence perception as an interactive parallel process. Science, 1975, 189, 226-228.

Marslen-Wilson, W. Linguistic descriptions and psychological assumptions in the study of sentence perception. In R.J. Wales and E.C.T. Walker (Eds.) New approaches to the study of language. Amsterdam, North-Holland, 1976.

Marslen-Wilson, W. and Tyler, L Processing structure of sentence perception. Nature, 1975, 257, 784-786.

Marslen-Wilson, W. and Tyler, L. Memory and levels of processing in a psycholinguistic context. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 112-119.

McKoon, G. and Ratcliff, R. Priming in episodic and semantic memory. Paper presented at the Psychonomic Society, St. Louis, November, 1976.

McKoon, G. and Ratcliff, R. The comprehension processes and memory structures involved in anaphoric reference. JVLVB, 1980, 19, 668-682.

O'Grady, W. A computational approach to pronominalization. Unpublished manuscript, University of Calgary, 1982.

Olson, J. and MacKay, D. Completion and verification of ambiguous sentences. JVLVB, 1974, 13, 457-470.

Perlmutter, D.M. and Soames, S. Syntactic argumentation and the structure of English. University of California Press, Los Angeles, 1979.

Posner, M. and Snyder, C. Attention and cognitive control. In R.L. Solso (Ed.), Information processing and cognition. Hillsdale, N.J.: Erlbaum, 1975.

Postal, P. Anaphoric islands. In Binnick, R., Davidson, G., Green, G. and Morgan, J. (Eds.) Papers from the Fifth Regional Meeting of the Chicago Linguistic Society. Chicago: The University of Chicago Press, 1969.

Postal, P. Pronominal epithets and similar items. Foundations of Language, 1972, 9(2).

Postal, P. On so-called "Pronouns" in English. (1966) Reprinted in D. Reibel and S. Schane (Eds.) Modern studies in English. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. A grammar of contemporary English. New York: Academic Press, 1972.

Ratcliff, R. and McKoon, G. Priming in item recognition: evidence for the propositional structure of sentences. JVLVB, 1978, 17, 403-417.

Rayner, K. Visual attention in reading: Eye movements reflect cognitive processes. Memory and Cognition, 1977, 4, 443-448.

Rayner, K. Eye movements in reading and information processing. Psychological Bulletin, 1978, 85, 618-660.

Reinhart, T. The syntactic domain of anaphora. Unpublished Ph.D. thesis, Massachusetts Institute of Technology, 1976.

Reinhart, T. Definite NP anaphora and c-command domains. Linguistic Inquiry, 1981, 12(4), 605-635.

Ross, J. On the cyclic nature of English pronominalization. In To Honor Roman Jakobson. The Hague: Mouton, 1967. Reprinted in D. Reibel and S. Schane (Eds.) Modern studies in English. Englewood Cliffs, N.J.: Prentice Hall, 1970.

Sachs, J. Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 1967, 2, 437-442.

Sachs, J. Memory in reading and listening to discourse. Memory and Cognition, 1974, 2, 95-100.

Sanford, A. and Garrod, S. Understanding written language: Explorations in comprehension beyond the sentence. Chichester, England: Wiley and Sons, 1981.

Schank, R. and Abelson, R. Scripts, plans, goals and understanding: An inquiry into human knowledge structure. Hillsdale, N.J.: Erlbaum, 1977.

Schlesinger, I. Sentence structure and the reading process. The Hague: Mouton, 1968.

Schmerling, S. A re-examination of "normal stress". Language, 1974, 50, 66-73.

Sheldon, A. The role of parallel function in the acquisition of relative clauses in English. JVLVB, 1974, 13, 272-281.

Shoben, E. Semantic and lexical decisions. In R. Puff (Ed.) Handbook of research methods in human memory and cognition. New York: Academic Press, 1982.

Smith, N. V. (Ed.) Mutual knowledge. New York: Academic Press, 1982.

Solan, L. Anaphora in child language. Unpublished Ph.D. thesis, Brandeis University, 1978.

Spencer, N. Differences between linguists and nonlinguists in intuitions of grammaticality-acceptability. Journal of Psycholinguistic Research, 1973, 12(2).

Sperber, D. and Wilson, D. in Smith, N.V. (Ed.) Mutual knowledge. New York: Academic Press, 1982.

Springston, J. Some cognitive aspects of presupposed coreferential anaphora. Unpublished Ph.D. thesis, Stanford University, 1975.

Sternberg, S. High-speed scanning in human memory. Science, 1966, 153, 652-654.

Sternberg, S. Two operations in character recognition: Some evidence from reaction-time measurements. Perception and Psychophysics, 1967, 2, 45-53.

Sternberg, S. The discovery of processing stages: Extensions of Donder's method. Attention and performance II. Acta Psychologica, 1969, 30, 276-315.

Stolz, W. Some experiments with queer sentences. Language and Speech, 1969, 12, 203-219.

Tanenhaus, M. and Seidenberg, M. Discourse context and sentence perception. Discourse Processes, 1981, 4, 197-220.

Tavakolian, S. Structural principles in the acquisition of complex sentences. Unpublished doctoral dissertation, U. Mass., 1977.

Taylor-Brown, K. Acquiring restrictions on forwards anaphora: A pilot study. Unpublished manuscript, University of Calgary.

Theios, J. Reaction time measurements in memory processes. In J. Bower (Ed.) The psychology of learning and motivation, vol. 7. New York: Academic Press, 1973.

Tyler, L. and Marslen-Wilson, W. The on-line effects of semantic context on syntactic processing. JVLVB, 1977, 16, 683-692.

Tyler, L. and Marslen-Wilson, W. Processing utterances in discourse contexts: On-line resolution of anaphors. Journal of Semantics, 1982, 1 (4), 297-314.

Tyler, L. Serial and interactive-parallel theories of sentence processing. Lingua, 1981, 49, 28-65.

Wasow, T. Anaphoric relations in English. Unpublished Ph.D. thesis, M.I.T., 1972.

Wasow, T. Anaphora in generative grammar. Brussels: E. Story-Scientia, 1979.

Waterhouse, V. Independent and dependent sentences. IJAL, 1963, 29, 45-54.

Webber, B. A formal approach to discourse anaphora. New York: Garland, 1979.

Webber, B. Discourse model synthesis: preliminaries to reference. In Webber, Sag, and Joshi (1981), 283-299.

Winer, B. Statistical principles in experimental design. New York: McGraw, 1971, 2nd edition.

Yekovitch, F. and Walker, C. Identifying and using referents in sentence comprehension. JVLVB, 1978, 17, 265-277.

Zegarra, L. and Smyth, R. SPART: Stimulus Presentation and Reaction Time. Unpublished programme, Hotel Dieu Hospital Kingston, Canada.

Ziff, P. About ungrammaticalness. Mind, 1964, 73, 204-214.

APPENDIX A: ITEMS, EXPERIMENT 1, LIST 1

1. I'm willing to pay him $2000 for Jack's car.
   Jack could get $2000.

2. In Trudeau's riding, he is quite well respected.
   The people in that riding respect Trudeau.

3. In Fred's graduation picture, he looks like a gorilla.
   Fred looks like a gorilla.

4. To him, I spoke in Ben's office.
   It was Ben that I spoke to.

5. After she dropped the kids off at daycare, Anne went to Lester's.
   Lester took the kids to daycare.

6. Neither Joel nor Martha thinks that he can win this battle.
   They don't think Joel can win.

7. The girl reported that Bill had struck his father.
   Bill struck the girl's father.

8. He was hit on the head before the lecturer had a chance to say anything.
   The lecturer was hit on the head.

9. Because Margaret made a fool of herself, the party was unpleasant for Bill.
   Bill made a fool of himself.

10. Blair was impressed by Laura's ability to detect his lies.
    Laura could detect Blair's lies.

11. You can't talk to him about Ben's kids.
    Ben won't discuss them.

12. In John's apartment he smokes pot.
    John smokes pot.

13. In Carter's home town, he spends his most pleasant hours.
    Carter enjoys himself there.

14. Instead of scolding him, Joyce should have explained what Christopher had done wrong.
    Joyce scolded Christopher.

15. In George's living room he placed his new brass bed.
George put it there.

16. For her, Betty's neighbours threw a party.
The party was for Betty.

17. Beth showed him her new tricks in the lawyer's jacuzzi.
The lawyer got to see her tricks.

18. She ran for a doctor after Henry fell down some stairs.
Henry ran for the doctor.

19. Michael's problems, he won't talk about.
Michael won't talk about them.

20. When Anita pulled the chair out from behind Phil, he fell on the floor with a thud.
Phil fell on the floor.

21. How obnoxious to Ben's friends he is.
Ben is obnoxious to them.

22. The worst thing about Jill's brother Arnold is his whimpy voice.
Arnold has a whimpy voice.

23. When Lorraine finishes school, she promised Jim to go to Paris.
Lorraine promised.

24. Lisa doesn't know where he lives, but I think Stewart has a house on Lower Union.
She doesn't know where Stewart lives.

25. Near Dan, he saw a snake.
Dan saw it.

26. Whenever Rachel discusses this with her analyst, he gets impatient.
The analyst gets impatient.

27. For Eileen's husband, she would give her life.
Eileen would die for him.

28. When she left the office, Bob saw the receptionist smile.
The receptionist left the office.

29. In the town where Julie lives with Jack, he has to pay a special gun tax.
Julie pays the tax.

30. Beside her, Kate noticed a leprechaun.
It was beside Kate.

31. Selling Margaret's photographs to his father was the best idea Pierre ever had.
Margaret's father bought the pictures.

32. Bonnie took Swedish with Fred before he flew to Stockholm.
Fred went to Stockholm.

33. Natalie and Steve used to live with his parents.
They lived with Steve's parents.

34. Jean told Keith that she had fallen in love with a soldier.
Jean had fallen in love.

35. The chauffeur parked his car near the banker's house.
The car belonged to the banker.

36. Abbey's first husband said she was too thin.
He said that Abbey was too thin.

37. Warren thinks Tom is a fool and HE thinks Judy is a hopeless case.
Tom thinks Judy is hopeless.

38. She is riding a horse in Gary's picture of Jane.
Jane is riding the horse.

39. It bothered her that Marie had failed.
Marie was bothered by it.

40. With Rosa's peacock feather, she tickles people.
Rosa tickles them.

41. Fred and Irene are waiting for the woman who is going to hire him.
The woman will hire Irene.

42. When he entered the room, Mary saw a salesman smile.
A salesman entered the room.

43. I told him that Fred couldn't come.
I told Fred.

44. One of the members of the Academy seems to have told her that Meryl is up for an Oscar.
Meryl heard about it.

45. The beautician wrote out a cheque for her dentist.
The dentist wrote the cheque.

46. We fired him, since Bert's behaviour had become irrational.
    We fired Bert.

47. For Cathy's T.V. I'm willing to pay her at most $300.
    I'd give Cathy $300.

48. In Marilyn's box, she put her valuable bracelet.
    Marilyn put the bracelet away.

49. In Ben's film about Rosa, she finds a husband.
    Rosa finds a husband.

50. With Eleanor's new job she'll make a lot of money.
    Eleanor will make a lot of money.

51. The waitress laughed at the comedian's joke because he sounded like Joan Rivers.
    The waitress sounded like Joan Rivers.

52. They finally decided that he would have to report to the new vice-president, whether McIntosh liked it or not.
    McIntosh would have a new boss.

53. Believe it or not, he is still considered a genius in Reagan's home town.
    Reagan is considered a genius.

54. Marie's work, you can't talk to her about.
    Marie won't talk about it.

55. After he woke up, Ernie went home.
    Ernie woke up and left.

56. We told him and Sue that Jack was a fool.
    We told Jack he was a fool.

57. Ted's girlfriend often forgets to brush her teeth after she eats garlic.
    Ted's girlfriend sometimes eats garlic.

58. Sally left her suitcase near Mark's car.
    It was Mark's suitcase.

59. The gossip about him upset Rick.
    The gossip was about Rick.

60. In Fred's office he is an absolute dictator.
    Fred is a dictator.

61. When he finishes this course, Bill promised his father to get a job.
Bill is taking a course.

62. In Frank's picture of Evelyn, she found a scratch.
Evelyn found the scratch.

63. We realized that whether she wanted to or not, Anna would have to report the incident to the police.
We realized that Anna might not want to do it.

64. For her new glasses, how much did she pay?
She paid for her own glasses.

65. In Sylvia's new condo near the waterfront, she lifts weights.
Sylvia lifts weights.

66. Janet won't tell anyone who Marnie met at her party.
It was Janet's party.

67. Fond of Donna's car though she is, I like it even more.
Donna is quite fond of her car.

68. Sue met him when Harvey was a little boy.
Sue met Harvey.

69. Jocelyn telephoned the D.J. because he had kept her records.
Jocelyn had kept the records.

70. Nancy should complain to Fred because he's a lousy lover.
Nancy is a lousy lover.

71. According to Margo, she's a fabulous cook.
Margo praises her own cooking.

72. Bernard told Evelyn that he was sick.
Evelyn was sick.

73. The speed at which she was driving made John nervous.
John was driving.

74. In Betty's wedding picture, she hopes to look like a lady.
Betty hopes she'll look good.

75. Gary killed the ballerina after she told the police about his stolen cars.
Gary told the police about the cars.

APPENDIX B

EXPERIMENT 1: Grammaticality of Coreference GROUP I

Instructions to Subjects

First, please fill out the confidential subject information at the top of your response sheet. Is your sheet marked Group I?

In this experiment I want to study the inferences you draw when you hear certain kinds of sentences. You will hear 75 pairs of sentences, consisting of a Target sentence and an Implication. Your task is to decide whether or not the Target sentence might in fact imply the second sentence, to rate your confidence in that decision, and to state whether that was your interpretation when you first heard the Target sentence.

Here is an example to answer in the section marked "PRACTICE":

1. TARGET: In spite of his illness, the mechanic managed to see Fred.
   IMPLICATION: Fred was ill.

Could the Target sentence imply that Fred was ill? Listen again (REPEAT 1.) Mark your answer in the YES or NO column, then decide how confident you are in that answer by writing a number from 1 (very uncertain) to 5 (very certain) in the middle column. Finally, in the third column, indicate whether that was the first meaning that came to mind when you heard the Target: mark your choice in the YES or NO column. Notice that if you said that the second statement was not a possible implication of the Target, you will have to answer NO here as well. During the test items, check the column headings from time to time to be sure you are marking the YES's and NO's in the right place.

When listening to the sentences, imagine that they are something you would hear in an ordinary conversation. You are to assume that the speaker is not trying to trick you or lie to you, and that he or she has not made an error of any kind. What I want to know is whether the person could expect you to interpret the Target sentence in the way that is suggested by the other sentence.

Do you have any questions before we go on to the other practice items?

Group I

2. TARGET: After he won the contest, Albert was interviewed by a CBC reporter.

   IMPLICATION: The reporter won the contest.

3. TARGET: The electrician noticed that George spilled the oil on him.

   IMPLICATION: The oil spilled on George.

4. TARGET: Joe told Frank that he liked Sue.

   IMPLICATION: Joe liked Sue.

5. TARGET: Ask Judy to tell the plumber not to leave her dog out.

   IMPLICATION: Judy has a dog.

Let's review these items. Did you find that your confidence was higher on some items than on others?

There are no "right" answers for these items: the purpose of this experiment is really to find out how people like you vary in your understanding of these items. This information may be used for later studies of the acquisition of language in children with communication disorders, as well as in normal children.

APPENDIX C

Experiment 1 Group 1: Response sheet

Name: Age:

Sex: Native language:

Highest year of schooling:

PRACTICE

	COULD the first sentence imply the second?		Confidence very uncertain 1 2 3 4 5 very certain	Was that your first interpretation?	
	YES	NO		YES	NO
1.	___	___	___	___	___
2.	___	___	___	___	___
3.	___	___	___	___	___
4.	___	___	___	___	___
5.	___	___	___	___	___

TEST

1.	___	___	___	___	___
2.	___	___	___	___	___
3.	___	___	___	___	___
4.	___	___	___	___	___
5.	___	___	___	___	___
6.	___	___	___	___	___
7.	___	___	___	___	___
8.	___	___	___	___	___
9.	___	___	___	___	___
10.	___	___	___	___	___
11.	___	___	___	___	___
12.	___	___	___	___	___

Group I

	COULD the first sentence imply the second?		Confidence very uncertain 1 2 3 4 5 very certain	Was that your first interpretation?	
	YES	NO		YES	NO
13.	___	___	___	___	___
14.	___	___	___	___	___
15.	___	___	___	___	___
16.	___	___	___	___	___
17.	___	___	___	___	___
18.	___	___	___	___	___
19.	___	___	___	___	___
20.	___	___	___	___	___
21.	___	___	___	___	___
22.	___	___	___	___	___
23.	___	___	___	___	___
24.	___	___	___	___	___
25.	___	___	___	___	___
26.	___	___	___	___	___
27.	___	___	___	___	___
28.	___	___	___	___	___
29.	___	___	___	___	___
30.	___	___	___	___	___
31.	___	___	___	___	___
32.	___	___	___	___	___
33.	___	___	___	___	___
34.	___	___	___	___	___
35.	___	___	___	___	___
36.	___	___	___	___	___
37.	___	___	___	___	___
38.	___	___	___	___	___
39.	___	___	___	___	___
40.	___	___	___	___	___
41.	___	___	___	___	___
42.	___	___	___	___	___

Group I

	COULD the first sentence imply the second?		Confidence very uncertain 1 2 3 4 5 very certain	Was that your first interpretation?	
	YES	NO		YES	NO
43.	___	___	___	___	___
44.	___	___	___	___	___
45.	___	___	___	___	___
46.	___	___	___	___	___
47.	___	___	___	___	___
48.	___	___	___	___	___
49.	___	___	___	___	___
50.	___	___	___	___	___
51.	___	___	___	___	___
52.	___	___	___	___	___
53.	___	___	___	___	___
54.	___	___	___	___	___
55.	___	___	___	___	___
56.	___	___	___	___	___
57.	___	___	___	___	___
58.	___	___	___	___	___
59.	___	___	___	___	___
60.	___	___	___	___	___
61.	___	___	___	___	___
62.	___	___	___	___	___
63.	___	___	___	___	___
64.	___	___	___	___	___
65.	___	___	___	___	___
66.	___	___	___	___	___
67.	___	___	___	___	___
68.	___	___	___	___	___
69.	___	___	___	___	___
70.	___	___	___	___	___
71.	___	___	___	___	___
72.	___	___	___	___	___
73.	___	___	___	___	___
74.	___	___	___	___	___
75.					

APPENDIX D: Master List

Experiment 2

```
 980  DATA THE CHILD WAS SCREAMIN
      G LOUDLY.
 990  DATA HE WAS STARING STRAIGH
      T AHEAD.
1000  DATA THE WAITER SAW A SNAK
      E IN FRONT OF HIM.
1010  DATA IT WAS IN FRONT OF TH
      E CHILD.
1020  DATA THE EVANGELIST WAS SP
      EAKING SOFTLY.
1030  DATA HE LOOKED TIRED.
1040  DATA THE MONK SAW A HALO O
      VER HIM.
1050  DATA THE HALO WAS OVER THE
      MONK.
1060  DATA THE ARTIST WAS AFRAID
      OF FLYING.
1070  DATA SHE WAS GOING TO BE S
      ICK.
1080  DATA THE NIECE PUT A BAG I
      N FRONT OF HER.
1090  DATA THE BAG WAS IN FRONT
      OF THE ARTIST.
1100  DATA THE CANDIDATE WAS UNE
      THICAL.
1110  DATA HE COULDN'T BE TRUSTE
      D.
1120  DATA THE CONGRESSMAN DROPP
      ED SOME COCAINE NEXT TO
      HIM.
1130  DATA THE COCAINE WAS NEXT
      TO THE CONGRESSMAN.
1140  DATA THE RESTAURANT OWNER
      WAS SHOUTING AND SCREAMIN
      G.
1150  DATA SHE WAS VERY DRUNK.
1160  DATA THE DAUGHTER SAW THE
      BOTTLE BEHIND HER.
1170  DATA THE BOTTLE WAS BEHIND
      THE DAUGHTER.
1180  DATA THE OFFICER WAS READY
      TO FIGHT.
1190  DATA HE LOOKED AROUND QUIC
      KLY.
1200  DATA THE HUSTLER SAW SOMET
      HING BEHIND HIM.
1210  DATA IT WAS BEHIND THE OFF
      ICER.
1220  DATA THE PROGRAMMER WAS MO
```

```
      RING FURIOUSLY.
1230  DATA  HE HAD TO BE FINISHED
      BY 8 O'CLOCK.
1240  DATA  THE DELIVERY BOY SAW
      THE TERMINAL IN      FRONT OF
      HIM.
1250  DATA  THE TERMINAL WAS IN F
      RONT OF THE          DELIVERY
      BOY.
1260  DATA  THE MERCHANT WAS SHOP
      PING ALONE.
1270  DATA  HE BROWSED FOR A WHIL
      E.
1280  DATA  THE SALESMAN PUT A BU
      Y DOWN IN FRONT      OF HIM.
1290  DATA  THE BOY WAS IN FRONT
      OF THE MERCHANT.
1300  DATA  THE DANCER WAS ON STA
      GE.
1310  DATA  SHE WAS DANCING BADLY
      .
1320  DATA  THE BUSINESSMEN SAW A
      LIGHT ABOVE HER.
1330  DATA  THE LIGHT WAS ABOVE T
      HE DANCER.
1340  DATA  THE TEACHER WAS FLYIN
      G TO TORONTO.
1350  DATA  SHE LOOKED TERRIFIED.

1360  DATA  THE STEWARDS SAW A MI
      SSILE BEYOND HER.
1370  DATA  THE MISSILE WAS BEYON
      D THE STEWARDS.
1380  DATA  THE OPERATOR WAS COOK
      ING IN THE NUDE.
1390  DATA  SHE REALLY DIDN'T CAR
      E ABOUT BEING SEEN.
1400  DATA  THE BACHELORS SAW A S
      HADOW BESIDE HER.
1410  DATA  THE SHADOW WAS BESIDE
       THE OPERATOR.
1420  DATA  THE JOURNALIST WAS GE
      TTING SCARED.
1430  DATA  HE WAS ALSO FREEZING.

1440  DATA  THE FARM GIRLS SAW A
      BLANKET OVER HIM.
1450  DATA  THE BLANKET WAS OVER
      THE FARM GIRLS.
1460  DATA  THE PSYCHOLOGIST WAS
      A BIT NERVOUS.
1465  DATA  SHE STEPPED FORWARD S
      LOWLY.
1470  DATA  THE WARRIORS SAW THE
      GOLD IN FRONT        OF HER.
1480  DATA  THE GOLD WAS IN FRONT
       OF THE WARRIORS.
1490  DATA  THE CLIENT WAS WAVING
      MADLY.
1500  DATA  HE SEEMED TO BE IN A
      HURRY.
1510  DATA  THE CALLGIRLS SAW A T
      AXI BEHIND HIM.
1520  DATA  THE TAXI WAS BEHIND T
      HE CLIENT.
1530  DATA  THE GUEST ENTERED SIL
```

```
        BUTLER.
1540  DATA  SHE COULDN'T SEE VERY
        WELL.
1550  DATA  THE USHERS NOTICED A
       POODLE BESIDE HER.
1560  DATA  THE POODLE WAS BESIDE
        THE USHERS.
1570  DATA  THE CELEBRITY HOPED H
       E WOULDN'T BE RECOGNIZED.
1580  DATA  HE WAS WORRIED BUT ST
       ILL EXCITED.
1590  DATA  THE STRIPPERS NOTICED
        A CAMERA          BEHIND H
       IM.
1600  DATA  THE CAMERA WAS BEHIND
        THE CELEBRITY.
1610  DATA  THE ALIEN STEPPED FOR
       WARD AND SMILED.
1620  DATA  SHE LOOKED QUITE UNUS
       UAL.
1630  DATA  IN FRONT OF HER THE G
       IRL GUIDE TOSSED    A GRENAD
       E.
1640  DATA  THE GRENADE WAS IN FR
       ONT OF THE ALIEN.
1650  DATA  THE FLORIST WAS WORKI
       NG DEFTLY.
1660  DATA  HE WANTED TO FINISH O
       N TIME.
1670  DATA  NEAR HIM THE HOST POS
        ITIONED A VASE.
1680  DATA  THE VASE WAS NEAR THE
        HOST.
1690  DATA  THE CLUB MEMBER WAS D
       RIVING TOO FAST.
1700  DATA  SHE FORGOT TO STOP BE
       FORE TURNING.
1710  DATA  IN FRONT OF HER THE B
       AG LADY SAW A CAT.
1720  DATA  THE CAT WAS IN FRONT
       OF THE MEMBER.
1730  DATA  THE SECRET AGENT WAS
       UPSET.
1740  DATA  SHE THOUGHT SHE WAS B
       EING FOLLOWED.
1750  DATA  BEHIND HER THE WIDOW
       SAW THE SIGNAL.
1760  DATA  THE SIGNAL WAS BEHIND
        THE WIDOW.
1770  DATA  THE HAIRDRESSER WAS S
       HAVING BADLY.
1780  DATA  HE WAS REALLY BURNED
       OUT.
1790  DATA  NEAR HIM THE ACTOR NO
       TICED SOME PILLS.
1800  DATA  THE PILLS WERE NEAR T
       HE ACTOR.
1810  DATA  THE THERAPIST GOT HOM
       E EARLY.
1820  DATA  HE LIKED TO CAUSE A S
       TIR.
1830  DATA  BEHIND HIM THE BUTLER
       SAW A FUNNY HAT.
1840  DATA  THE HAT WAS BEHIND TH
       E THERAPIST.
```

```
1850 DATA THE VIOLINIST HAD PRA
     CTICED FOR WEEKS.
1860 DATA HE WAS NERVOUS BEFORE
     THE CONCERT.
1870 DATA OVER HIM THE TENOR SA
     W A SPOTLIGHT.
1880 DATA THE LIGHT WAS OVER TH
     E TENOR.
1890 DATA THE FASHION PHOTOGRAP
     HER WAS WORKING      OUTDOORS
     .
1900 DATA SHE SQUINTED AS SHE F
     OCUSSED.
1910 DATA NEAR HER THE BRIDESMA
     ID SPOTTED A SKUNK.
1920 DATA THE SKUNK WAS NEAR TH
     E PHOTOGRAPHER.
1930 DATA THE INSTRUCTOR LOOKED
      UP.
1940 DATA SHE WAS DISGUSTED AND
      THOUGHT SHE MIGHT  VOMIT.
1950 DATA ABOVE HER THE BARMEN
     SAW A PIGEON.
1960 DATA THE PIGEON WAS ABOVE
     THE INSTRUCTOR.
1970 DATA THE SINGER WAS RESOUN
     DINGLY SUCCESSFUL.
1980 DATA HE WAS HAPPIER THAN H
     E'D EVER BEEN.
1990 DATA IN FRONT OF HIM THE M
     OTHERS THREW FLOWERS.
2000 DATA THE FLOWERS WERE IN F
     RONT OF THE MOTHERS
2010 DATA THE LIFEGUARD STOOD U
     P AND DIVED.
2020 DATA SHE PLUNGED STRAIGHT
     DOWN.
2030 DATA UNDER HER THE PLAYBOY
     S SPOTTED A WATCH.
2040 DATA THE WATCH WAS UNDER T
     HE LIFEGUARD.
2050 DATA THE SERVANT WAS READY
     .
2060 DATA HE SMILED SLIGHTLY.
2070 DATA UNDER HIM THE GIRLS H
     AD PLACED A FIRE-   CRACKER.

2080 DATA THE FIRECRACKER WAS U
     NDER THE GIRLS.
2090 DATA THE POET SAT DOWN WEA
     VILY.
2100 DATA SHE FELT HOT AND STIC
     KY.
2110 DATA NEAR HER THE GENTLEME
     N SPIED A WALLET.
2120 DATA THE WALLET WAS NEAR T
     HE GENTLEMEN.
2130 DATA THE HISTORIAN WAS TIM
```

```
E TRAVELLING.
2140  DATA  HE FELT CONFUSED AND
      DISORIENTED.
2150  DATA  UNDER HIM THE MAIDENS
      SAW A BLACK HOLE.
2160  DATA  THE BLACK HOLE WAS UN
      DER THE HISTORIAN.
2170  DATA  THE CLEANER WAS ALWAY
      S LAUGHING.
2180  DATA  SHE LOVED TO JOKE ARO
      UND.
2190  DATA  UNDER HER THE CHOIR B
      OYS PLACED A MOUSE.
2200  DATA  THE MOUSE WAS UNDER T
      HE CHOIR BOYS.
2210  DATA  THE MODEL WAS DANCING
       HAPPILY.
2220  DATA  SHE WAS THRILLED TO H
      AVE BEEN INVITED.
2230  DATA  NEAR HER THE DISCOUNT
      S NOTICED A THRONE.
2240  DATA  THE THRONE WAS NEAR T
      HE MODEL.
2250  DATA  THE CITIZEN WAS WAITI
      NG EXPECTANTLY.
2260  DATA  HE WAS SURE HE WOULD
      BE RICH.
2270  DATA  BESIDE THE PRINCE HE
      HAD FOUND A DIAMOND
2280  DATA  THE CITIZEN HAD FOUND
       A DIAMOND.
2290  DATA  THE CLERK WAS NO LONG
      ER HUNGRY.
2300  DATA  HE DID NOTHING BUT CO
      MPLAIN.
2310  DATA  NEXT TO THE WAITER HE
       HAD SEEN A BUG.
2320  DATA  THE WAITER SAW A BUG.

2330  DATA  THE PATIENT WAS UNCOM
      FORTABLE.
2340  DATA  SHE COULD BARELY MOVE
      .
2350  DATA  BEHIND THE GIRL SHE C
      OULD SEE A WHEEL-  CHAIR.
2360  DATA  THE PATIENT COULD SEE
       A WHEELCHAIR.
2370  DATA  THE ORGANIZER WAS DIS
      SATISFIED.
2380  DATA  HE HAD TO RECONSIDER.

2390  DATA  ABOVE THE ERRAND BOY
      HE SPIED A CALENDAR.
2400  DATA  THE ERRAND BOY SPIED
      A CALENDAR.
2410  DATA  THE CYCLIST DISMOUNTE
      D.
2420  DATA  SHE WAS SWEATING PROF
      USELY.
2430  DATA  BESIDE THE SECRETARY
      SHE NOTICED A       FOUNTAIN
      .
2435  DATA  THE SECRETARY NOTICED
       A FOUNTAIN.
2440  DATA  THE ORGANIST PLAYED P
      OORLY.
```

```
2450  DATA  SHE JUST COULDN'T PER
      FORM.
2460  DATA  BEHIND THE WAITRESS S
      HE SAW A FLASH.
2470  DATA  THE OPTOMIST SAW A FL
      ASH.
2480  DATA  THE DEALER WAS BIDDIN
      G TOO HIGH.
2490  DATA  SHE WANTED TO WIN.
2500  DATA  BEHIND THE TALL WOMAN
       SHE SAW A          CANDELAB
      RA.
2510  DATA  THE TALL WOMAN SAW TH
      E CANDELABRA.
2520  DATA  THE OFFICER WAS SUSPI
      CIOUS.
2530  DATA  HE WAS GOING TO CHECK
       CAREFULLY.
2540  DATA  BEHIND THE BARMAN HE
      SPOTTED THE LICENCE.
2550  DATA  THE OFFICER SPOTTED T
      HE LICENCE.
2560  DATA  THE STAR SMILED BRIGH
      TLY.
2570  DATA, THEN SHE WAVED.
2580  DATA  AHEAD OF THE SAILORS
      SHE SAW A LIMOUSINE
2590  DATA  THE STAR SAW THE LIMO
      USINE.
2600  DATA  THE FLYER TURNED.
2610  DATA  THEN HE STARTED COUGH
      ING.
2620  DATA  AROUND THE WIVES HE S
      AW SMOKE.
2630  DATA  THE WIVES SAW THE SMO
      KE.
2640  DATA  THE STUDENT WAS STUDY
      ING TOO HARD.
2650  DATA  SHE NEEDED TO RELAX F
      OR A WHILE.
2660  DATA  BESIDE THE LITTLE BOY
      S SHE NOTICED A COT
2670  DATA  THE STUDENT NOTICED A
       COT.
2680  DATA  THE MIME ARTIST WAS Q
      UITE ENTERTAINING.
2690  DATA  HE WAS VERY EXPERIENC
      ED.
2700  DATA  NEAR THE GIRLS HE SAW
       THE COSTUMES.
2710  DATA  THE GIRLS SAW THE COS
      TUMES.
2720  DATA  THE PAINTER WAS HALLU
      CINATING.
2730  DATA  SHE WAS BECOMING STRA
      NGER AND STRANGER.
2740  DATA  NEAR THE CATMEN SHE P
      AINTED A UNICORN.
2750  DATA  THE CATMEN PAINTED A
      UNICORN.
2760  DATA  THE DIET COUNSELLOR F
      ELT CHALLENGED.
2770  DATA  HE HAD TO KEEP ALERT.

2780  DATA  IN FRONT OF THE FAT W
      OMEN HE SPOTTED THE BROWNIES
```

```
2785 DATA THE COUNSELLOR SPOTTE
     D THE BROWNIES.
2790 DATA THE SPECTATOR WANTED
     TO PHONE HOME
2800 DATA BUT SHE WAS GOING TO
     BE INTRODUCED.
2810 DATA NEAR THE BULLFIGHTERS
      SHE SAW A          TELEPHON
     E.
2820 DATA THE BULLFIGHTERS SAW
     A TELEPHONE.
2830 DATA THE AUTHOR SLEPT SOUN
     DLY.
2840 DATA HE WAS DREAMING IN TE
     CHNICOLOUR.
2850 DATA NEXT TO THE HEROINES
     HE PERCEIVED A      CASTLE.
2860 DATA THE AUTHOR PERCEIVED
     A CASTLE.
2870 DATA THE VISITOR STROLLED
     AROUND ALL DAY.
2880 DATA SHE LOVED BEING AWAY.

2890 DATA SHE HEARD A NOISE NEA
     R THE COMPEL.
2900 DATA THE VISITOR HEARD THE
     NOISE.
2910 DATA THE OWNER WAS PERTURB
     ED.
2920 DATA HE DIDN'T KNOW WHAT T
     O DO.
2930 DATA HE SPOTTED A GUN BESI
     DE THE SAILOR.
2940 DATA THE SAILOR SPOTTED TH
     E GUN.
2950 DATA THE SWIMMER WAS EXHAU
     STED.
2960 DATA SHE WAS AFRAID SHE WO
     ULD PASS OUT.
2970 DATA SHE CAUGHT SIGHT OF A
      BOAT BEYOND THE    MERMAID.

2980 DATA THE SWIMMER SAW THE B
     OAT.
2990 DATA THE MUSIC DIRECTOR WA
     S NOT AMUSED.
3000 DATA HE WAS VERY SERIOUS.
3010 DATA HE NOTICED A TOY BESI
     DE THE CHOIR BOY.
3015 DATA THE CHOIR BOY NOTICED
     THE TOY.
3020 DATA THE SLAVE HAD TO ESCA
     PE.
3030 DATA HE WAS AFRAID HE WOUL
     D BE KILLED.
3040 DATA HE SAW A KNIFE BEHIND
     THE MASTER.
3050 DATA THE MASTER SAW THE KN
     IFE.
3060 DATA THE GARDENER WAS SUPE
     RSTITIOUS.
3070 DATA SHE WAS THOUGHT TO BE
     PSYCHIC.
3080 DATA SHE PERCEIVED A GREEN
     GLOW OVER          THE NITE
```

```
H.
3090  DATA  THE GARDENER PERCEIVE
      D THE GLOW.
3100  DATA  THE DEMONSTRATOR WAS
      ANXIOUS TO SUCCEED.
3200  DATA  SHE WANTED TO PROVE T
      HAT SHE WAS RIGHT.
3210  DATA  SHE HAD PLANTED A CAM
      ERA BEHIND THE       PORNO QU
      EEN.
3220  DATA  THE PORNO QUEEN PLANT
      ED THE CAMERA.
3230  DATA  THE MONARCHIST WAS EC
      STATIC.
3240  DATA  HE WAS BEAMING WITH P
      LEASURE.
3250  DATA  HE COULD SEE THE JEWE
      LS ON THE KING.
3260  DATA  THE MONARCHIST COULD
      SEE THE JEWELS.
3270  DATA  THE COOK WAS IN TROUB
      LE.
3280  DATA  SHE WAS AFRAID TO RUN
      AWAY.
3290  DATA  SHE COULD SEE THE E I
      T BEHIND THE         POLICEME
      N.
3300  DATA  THE COOK COULD SEE TH
      E EXIT.
3310  DATA  THE DIVER WAS EMBARRA
      SSED.
3320  DATA  HE DIDN'T REALIZE WHE
      RE HE WAS.
3330  DATA  HE SPOTTED A DOOR NE
      T TO THE HOOKERS.
3340  DATA  THE HOOKERS SPOTTED T
      HE DOOR.
3350  DATA  THE TOURIST WAS GLAD
      TO BE ABLE TO SIT   DOWN.
3360  DATA  SHE NEEDED TO REST AF
      TER WALKING ALL DAY.
3370  DATA  SHE SAW A FAT NEAR TH
      E FISHERMEN.
3380  DATA  THE TOURIST SAW A FAT
      .
3390  DATA  THE SUPERVISOR DIDN'T
      SPEAK.
3400  DATA  HE JUST SIGHED SOFTLY
      .
3410  DATA  HE HAD FOUND THE REPO
      RT BESIDE THE       OFFICE G
      IRLS.
3420  DATA  THE OFFICE GIRLS FOUN
      D THE REPORT.
3430  DATA  THE RESEARCHER WAS NE
      ARLY FINISHED.
3440  DATA  SHE HAD DONE ALL SHE
      COULD.
3450  DATA  SHE NOTICED A TELEPHO
      NE NEAR THE BOXERS.
3460  DATA  THE BOXERS NOTICED A
      TELEPHONE.
3461  DATA  THE EMPLOYEE WAS QUIT
      TING.
3462  DATA  SHE WASN'T EARNING EN
      OUGH.
```

```
3460  DATA  SHE SAW A TIP IN FRON
      T OF THE BUSBOYS.
3464  DATA  THE EMPLOYEE SAW THE
      TIP.
3470  DATA  THE PIANIST WAS REALL
      Y LOUSY.
3480  DATA  SHE WASN'T VERY POPUL
      AR.
3490  DATA  SHE SAW A CURTAIN IN
      FRONT OF THE        SOLDIERS
      .
3500  DATA  THE SOLDIERS SAW A CU
      RTAIN.
3510  DATA  THE GHOST WAS SHIMMER
      ING OVERHEAD.
3520  DATA  SHE WAS WEARING A LON
      G WHITE ROBE.
3530  DATA  SHE SAW A CANDLE NEXT
       TO THE SORCERERS.
3540  DATA  THE GHOST SAW A CANDL
      E.
```

APPENDIX E: SPART LISTING, EXPERIMENT 2

```
10  REM  CREATED JAN 10,1984
20  REM  BY LUIS ZEGARRA
30  REM  REACTION TIME
40  REM   MODIFIED JULY 29, 1984
45  REM  PART ONE OF TEST
50  DIM BLOCK$(64,4),NUM(64),BN(6
    4)
60  DIM TIME(64,4),A11$(64,4)
70  HOME
80  PRINT "PLEASE ENTER THE FOLLO
    WING"
90  PRINT "IN ONE LINE (NO SPACES
    )"
100  PRINT : PRINT "INITIALS (2 L
    ETTERS)"
110  PRINT "AGE (2 DIGITS)"
120  PRINT "SEX (M OR F)"
121  PRINT "SUBJECT NUMBER"
122  PRINT "LIST NUMBER"
130  INPUT CODE$
140  HOME : VTAB 15
150  HTAB 13
160  PRINT "PLEASE WAIT"
170  REM  READ SENTENCES\ INTO BLO
    CKS
190  FOR J = 1 TO 64
190  FOR I = 1 TO 4
200  READ BLOCK$(J,I)
210  NEXT I
215  READ L$: READ L$: READ L$: READ
    L$
220  NEXT J
230  REM  RANDOMIZE SEQUENCE OF P
    RESENTATION
240  LET M = 0
250  FOR I = 1 TO 64: LET NUM(I) =
    0: NEXT I
260  LET N =  INT (64 *  RND (1))
     + 1
270  IF NUM(N) <  > 0 THEN  GOTO
    260
280  LET NUM(N) = 1
290  LET M = M + 1
300  LET BN(M) = N
310  IF M <  > 64 THEN  GOTO 260
320  REM  DISPLAY SENTENCES AND T
    IME\RESPONSE
330  HOME
340  FOR M = 1 TO 64
350  LET J = BN(M)
360  HOME
370  VTAB 10: HTAB 7
380  PRINT "    *** RELAX ***
390 X =  PEEK (49152)
```

```
400  IF X < 128 THEN  GOTO 390
410  GET B$
420  FOR I = 1 TO 4: HOME
421  IF I = 3 THEN  FOR D = 1 TO
     500: NEXT D
422  IF I = 4 THEN  FOR D = 1 TO
     250: NEXT D
430  VTAB 15
440  IF I = 4 THEN  INVERSE
450  PRINT BLOCK$(J,I)
460  NORMAL
471  IF I <  > 2 THEN  GOTO 480
472  PRINT "..................
     ..........;......."
473  VTAB 23
480  & CLRTIMER
490 X =  PEEK (49152)
500  IF X < 128 THEN  GOTO 490
510  & TIMERIN,(T%) = 0
520 Z =  PEEK (49168)
530 X = X - 128
540 A$ =  CHR$ (X)
550  REM  21 IS MEASURED PROCESSO
     R S EXECUTION TIME
560 TIME(J,I) = (0 - 21)
570  IF A$ = "0" THEN  GOTO 610
580  IF A$ = "1" THEN  GOTO 630
590  LET A11$(J,I) = " "
600  GOTO 640
610  LET A11$(J,I) = "0"
620  GOTO 640
630  LET A11$(J,I) = "1"
640  IF TIME(J,I) > 10000 THEN  LET
     TIME(J,I) =  - 999
650  NEXT I
660  NEXT M
670  HOME
680  VTAB 13
690  HTAB 15
700  FLASH : PRINT "THANK YOU": NORMAL

740 D$ =  CHR$ (4)
745  ONERR  GOTO 940
750  PRINT D$;"OPEN"CODE$",P1,D2"

760  PRINT D$;"WRITE"CODE$",P1"
764  FOR J = 1 TO 64
765 M = J * 2 - 1
767  FOR I = 1 TO 4
770  PRINT M","I;TIME(J,I);A11$(J
     ,I)
780  NEXT I
790  NEXT J
800  PRINT D$;"CLOSE"CODE$",P1"
810  POKE 43624,1
820 I$ =  CHR$ (9)
830  PR# 1
840  PRINT I$;"100N"
850  PRINT CODE$: PRINT : PRINT
860  FOR J = 1 TO 64
865 M = J * 2 - 1
870  FOR I = 1 TO 4
880  PRINT M","I,TIME(J,I)"SECOND
     S",A11$(J,I)
890  NEXT I
900  PRINT : PRINT
```

```
910 ;NEXT J
920  PR# 0
930  GOTO 965
940 Y =  PEEK (222)
941  IF  NOT (Y = 4 OR Y = 8 OR Y
     = 9) THEN  GOTO 750
943  PRINT  CHR$ (7): HOME
945  VTAB 13: HTAB 5
946  IF Y = 8 THEN  PRINT "*** CL
     OSE DOOR OR INIT DISK ***"
950  IF Y = 4 THEN  PRINT "*** DI
     SK WRITE PROTECTED ***"
953  IF Y < > 9 GOTO 961
955  PRINT D$;"DELETE"CODE$".P1"
958  PRINT "***DISK FULL - CHANGE
      DISK***"
961  PRINT : PRINT : PRINT :"PRES
     S RETURN TO CONTINUE"
963  GET Y$: PRINT
964  RESUME
965  END
```

APPENDIX F:

Instructions: Experiment 2

In this experiment I am examining how people are able to understand sentences when they are presented together as a connected paragraph. You will be reading three-sentence paragraphs on the computer screen. Each is followed by a true/false question. You will be able to control the rate at which the sentences appear by pressing the space bar (marked "NEXT SENTENCE"), and your responses to the questions will be timed.

Let me demonstrate with one of the 15 practice items you will try. We begin with a "RELAX" signal. Whenever you see this, you may pause for as long as you wish. Now we can look at the first two sentences, the CONTEXT sentences, by pressing the space bar. Notice that when the second sentence appears the first is erased - you cannot go back to see it again. Spend as much time as you need to understand these first two sentences. They provide the context for your understanding of the TARGET sentence. You know the Target is coming by the row of dots after context sentence number 2. Hit the space bar again. The screen will go blank for a second. When the Target sentence comes up, read it normally, and hit the bar again the instant you feel that you have fully understood it. Don't try to read as fast as possible, and don't try to learn the sentence by heart. I want to measure how long people take to understand different written materials.

Finally, your comprehension will be tested by a true/false question. Hit the correct button, without looking down at your fingers, as soon as you possibly can, without making any mistakes. I am giving cash prizes of $25, $15 and $5 for the best performers on this task.

APPENDIX G: MASTER LIST, EXPERIMENT 3

```
1000 DATA DENNIS AND LAURA STEP
     PED OUTSIDE AND HE LOCKED T
     HE DOOR.
1005 DATA DENNIS
1010 DATA THEY HAD BEEN INSIDE.

1015 DATA SUE AND BOB FORGOT TH
     E ADDRESS BUT SHE REMEMBER
     ED THE PHONE NUMBER.
1020 DATA SUE
1025 DATA THEY REMEMBERED THE A
     DDRESS.
1030 DATA BILL AND ANNE STUDIED
     COOKING WHEN HE WORKED I
     N TORONTO.
1035 DATA BILL
1040 DATA THEY STUDIED COOKING.

1045 DATA JANE AND JIM ARE BOTH
     GOOD PARENTS BUT SHE TAKE
     S CARE OF THE KIDS.
1050 DATA JANE
1055 DATA JIM TAKES CARE OF THE
     KIDS.
1060 DATA DICK AND BARB HATED T
     HE WINTERS SO HE BOUGHT A
     HOUSE IN FLORIDA.
1065 DATA DICK
1070 DATA THEY HATED THE WINTER
     S.
1075 DATA KATE AND MIKE PLANTED
     A GARDEN AND SHE CANNED T
     HE TOMATOES.
1080 DATA KATE
1085 DATA THEY DIDN'T PLANT A G
     ARDEN.
1086 DATA JOYCE AND DAVE LIVE I
     N VERMONT SO HE HAS TO DRIVE
     QUITE FAR.
1087 DATA DAVE
1088 DATA THEY LIVE IN MINNESOT
     A.
1090 DATA JACK AND JOAN ARE ALW
     AYS FIGHTING SO SHE WON'T DI
     SCUSS MARRIAGE ANYMORE.
1095 DATA JOAN
1100 DATA THEY NEVER FIGHT.
1105 DATA NANCY AND HARRY CAME
     TO THE COUNTRY AND HE WENT
     ON A HIKE.
1110 DATA HARRY
1115 DATA HARRY STAYED IN TOWN.

1120 DATA STEVE AND JEAN ARGUED
     ABOUT MONEY WHEN SHE GOT
     THE BILL.
1125 DATA JEAN
```

```
1130  DATA  A BILL ARRIVED.
1135  DATA  BETTY AND LARRY LIVED
      IN BELLEVILLE BUT  HE HATED
      THE PLACE.
1140  DATA  LARRY
1142  DATA  THEY LIVED IN TRENTON
      :
1145  DATA  RICHARD AND SUSAN BOU
      GHT A FARM WHEN SHE MOVED BA
      CK FROM ALBERTA.
1150  DATA  SUSAN
1155  DATA  THEY BOUGHT A FARM.
1175  DATA  CHARLES AND ALICE PEN
      TED A CAR AND SHE  DROVE IT
       TO OTTAWA.
1180  DATA  CHARLES
1185  DATA  THE CAR WENT TO OTTAW
      A.
1190  DATA  ELLEN AND JERRY CLEAN
      ED THE GARAGE BUT  HE DIDN'
      T CUT THE GRASS.
1195  DATA  ELLEN
1200  DATA  NOBODY CLEANED THE GA
      RAGE.
1205  DATA  RICK AND RUTH WERE CO
      VERED BY OHIP SO .  SHE DIDN
      'T WORRY ABOUT THE BILLS.
1210  DATA  RICK
1215  DATA  THEY WORKED FOR OHIP.

1220  DATA  HELEN AND ROBERT BECA
      ME CONSERVATIVES  WHEN HE
      LEFT THE LIBERAL PART.
1225  DATA  HELEN
1230  DATA  THEY BECAME COMMUNIST
      S.
1235  DATA  DAVID AND JANET ABAND
      ONED THE PROJECT  AND SHE
      GOT ANOTHER JOB.
1240  DATA  DAVID
1245  DATA  THEY ABANDONED THE PR
      OJECT.
1250  DATA  JANE AND BRUCE ATE TH
      E PIZZA BUT HE GOT  SICK AFT
      ERWARDS.
1255  DATA  JANE
1256  DATA  THEY ATE LIVERWURST.
1257  DATA  DAN AND BETH REFINISH
      ED THE FURNITURE  WHEN HE
      SAW HOW BADLY IT WAS SCRATCH
      ED.
1258  DATA  BETH
1259  DATA  THEY BURNED THE FURNI
      TURE.
1265  DATA  GAIL AND CARL FOUNDED
       THE COMPANY SO SHE HAD SPEC
      IAL PRIVILEGES.
1270  DATA  CARL
1275  DATA  THEY FOUNDED A COMPAN
      Y.
1280  DATA  JAMES AND JILL BOUGHT
       SOME MAGAZINES  WHEN HE
      ENTERED THE HOSPITAL.
1285  DATA  JILL
1290  DATA  JAMES AND JILL SOLD T
      HE MAGAZINES
```

```
1295  DATA  JOANNE AND HENRY LIKE
      TRAVELLING SO SHE  BOUGHT S
      ONE TICKETS.
1300  DATA  HENRY
1305  DATA  THEY LIKE TRAVELLING.

1310  DATA  BRIAN AND BONNIE LIVE
      D IN MONTREAL AND  HE COMMU
      TED ON WEEKDAYS.
1315  DATA  BONNIE
1320  DATA  THEY LIVED IN VANCOUV
      ER.
1325  DATA  PAM AND RON WROTE SOM
      E ARTICLES BUT SHE  HAD ALL
      THE IDEAS.
1330  DATA  RON
1335  DATA  THEY WROTE SOME ARTIC
      LES.
1355  DATA  WILLIAM AND PEGGY TOO
      K DANCE LESSONS BUT I WAS TE
      RRIBLE AT IT.
1360  DATA  WILLIAM
1365  DATA  THEY TOOK DANCE LESSO
      NS.
1370  DATA  RITA AND ROGER UNDERS
      TAND FRENCH SO I    HAD NO T
      ROUBLE IN PARIS.
1375  DATA  RITA
1380  DATA  THEY DON'T SPEAK FREN
      CH.
1385  DATA  BARRY AND SANDRA WERE
       IRRITABLE WHEN I   GOT HOME
       FROM WORK.
1390  DATA  BARRY
1395  DATA  SANDRA WAS IRRITABLE.

1400  DATA  SHERRY AND BRIAN ORDE
      RED MARTINIS AND I  ASKED FO
      R SOME PEANUTS.
1405  DATA  SHERRY
1410  DATA  THEY ORDERED BEER.
1415  DATA  EDWARD AND EILEEN HAD
       A PROBLEM SO I     SOLVED I
      T.
1420  DATA  EDWARD
1425  DATA  I SOLVED THE PROBLEM.

1430  DATA  JULIE AND HOWARD BOTH
       GOT FIRED SO I     DECIDED
      TO SUE THE COMPANY.
1435  DATA  JULIE
1440  DATA  NOBODY GOT FIRED
1441  DATA  MARGE AND TOD WENT ON
       A DIET AND I       DECIDED
      TO QUIT SMOKING.
1442  DATA  TOD
1443  DATA  THEY WENT ON AN EATIN
      G BINGE.
1445  DATA  WARREN AND CONNIE RUN
      TED FOR BARGAINS    WHEN I N
      OTICED THE AUCTION.
1450  DATA  CONNIE
1455  DATA  THEY LOOKED FOR BARGA
      INS.
1460  DATA  BRIAN AND NEIL ENTERED
       THE CONTEST AND    I WON A
```

```
     PRIZE.
1465  DATA  NEIL
1470  DATA  THEY REFUSED TO ENTER
      THE CONTEST
1475  DATA  THOMAS AND CHERYL DES
     IGNED A COTTAGE BUT  I WAS U
     NABLE TO HAVE IT BUILT.
1480  DATA  CHERYL
1485  DATA  THEY DESIGNED A COTTA
     GE.
1490  DATA  BETSY AND KEVIN SLEPT
      UNTIL NOON SO I    WAS LATE
      FOR WORK.
1495  DATA  KEVIN
1500  DATA  THEY GOT UP AT 8 A.M.

1505  DATA  RUSSELL AND BRENDA WE
     RE WORRIED WHEN I   COULDN'T
      FIND THE TICKETS.
1510  DATA  BRENDA
1515  DATA  THEY WERE WORRIED.
2000  DATA  BERT FLEW TO JAPAN WI
     TH LUCY BECAUSE SHE HAD RELA
     TIVES THERE.
2010  DATA  EVELYN
2020  DATA  THEY WENT TO JAPAN.
2030  DATA  ALBERT AND FLORA WATC
     HED THE ELECTION   ON TELEV
     ISION.
2040  DATA  SAM
2050  DATA  THEY WATCHED THE HOCK
     EY GAME.
2060  DATA  TIMMIE SNATCHED THE G
     UM AWAY FROM       ALLISON
     SO I TOOK IT AWAY.
2070  DATA  MARIE
2080  DATA  I TOOK THE GUM AWAY.
2090  DATA  DONNA WENT TO THE CON
     VENT WITH HERBERT  BUT THEY
      DIDN'T SEE ANY NUNS.
```

```
2000  DATA  BERT FLEW TO JAPAN WI
      TH LUCY BECAUSE SHE HAD RELA
      TIVES THERE.
2010  DATA  EVELYN
2020  DATA  THEY WENT TO JAPAN.
2030  DATA  ALBERT AND FLORA WATC
      HED THE ELECTION  ON TELE
      ISION.
2040  DATA  SAM
2050  DATA  THEY WATCHED THE HOCK
      EY GAME.
2060  DATA  TINNIE SNATCHED THE G
      UN AWAY FROM  ALLISON
      SO I TOOK IT AWAY.
2070  DATA  MARIE
2080  DATA  I TOOK THE GUN AWAY.
2090  DATA  DONNA WENT TO THE CON
      VENT WITH HERBERT  BUT THEY
       DIDN'T SEE ANY NUNS.
2100  DATA  EILEEN
2110  DATA  THEY VISITED SOME NUN
      S.
2120  DATA  LUCY AND DOUG RAISED
      POODLES BUT HE  HATED TH
      E LITTLE BEASTS.
2130  DATA  GORDON
2140  DATA  DOUG HATED THE POODLE
      S.
2150  DATA  EVEN THOUGH THEY ELOP
      ED MARYANNE AND  NORMAN G
      OT WEDDING GIFTS.
2160  DATA  MARTIN
2170  DATA  SHE  DIDN'T GET GIFTS
      .
2180  DATA  WITH SEVEN KIDS ALIST
      AIR AND BETSY HERE  ALWAYS T
      IRED.
2190  DATA  HERMAN
2200  DATA  THEY HAD SEVEN CHILDR
      EN.
2210  DATA  FLO WANTED TO GET A D
      OG FOR DERRICK BUT  SHE HAD
      ALLERGIES.
2220  DATA  MAY
2230  DATA  DERRICK HAD ALLERGIES
      .
2231  DATA  CINDY AND TROY MOVED
      TO VANCOUVER WHEN  HE HEARD
       ABOUT THE WARRANT.
2232  DATA  RYAN
2233  DATA  THEY MOVED TO HALIFAX
      .
2240  DATA  RETIRING TO BERMUDA W
      ITH ARTHUR MEANT  THAT DEN
      ISE WOULD HAVE A NICE TAN.
```

```
2250  DATA  MAR,
2260  DATA  THEY RETIRED TO BERMU
      DA.
2270  DATA  MARILYN'S DIVORCE FRO
      M CECIL SURPRISED  EVERYONE
      .
2280  DATA  SUSIE
2290  DATA  NOBODY WAS SURPRISED.

2300  DATA  PETER AND JODIE SIGNE
      D THE CONTRACT BUT  SHE DIDN
      'T WANT TO DO IT.
2310  DATA  FRANK
2320  DATA  THEY SIGNED THE CONTR
      ACT.
2330  DATA  HEIDI AND HEINZ HAD S
      URGERY BUT HIS WAS  UNSUCCES
      SFUL.
2340  DATA  MALCOLM
2350  DATA  HIS SURGERY WAS SUCCE
      SSFUL.
2360  DATA  WHEN NIXON WAS PRESID
      ENT ADRIENNE AND  PHILIP W
      ORKED FOR THE C.I.A.
2370  DATA  AMY
2380  DATA  THEY WORKED FOR THE C
      .I.A.
2390  DATA  STEVIE AND JOYCE BUY
      RECORDS BUT SHE  NEVER PL
      AYS THEM.
2400  DATA  GARY
2410  DATA  SHE PLAYS THE RECORDS
      OFTEN.
2420  DATA  TIM AND HAZEL WENT TO
      LAS VEGAS BUT SHE  DIDN'T G
      AMBLE.
2430  DATA  DONALD
2440  DATA  SHE DIDN'T GAMBLE.
2450  DATA  SARAH AND SHELDON T,P
      ED THEIR PAPERS  WHEN SHE
      GOT BACK FROM CLASS.
2460  DATA  WROTE
2500  DATA  THEY TYPED THE PAPERS
      .
2510  DATA  CRAIG WAS FLORENCE S
      ACCOMPLICE IN THE  JEWEL TH
      EFT BUT HE WAS LATER CAUGHT.

2520  DATA  BOUGHT
2530  DATA  FLORENCE WAS CAUGHT.
2540  DATA  CHARLOTTE AND ERIC HA
      D AN AUDIENCE WITH  THE POPE
      SO I WAS IMPRESSED.
2550  DATA  PRIEST
2560  DATA  ERIC SAW THE POPE.
2570  DATA  RAVIOLI WAS WHAT ROD
      AND CORINNE ORDERED BUT I GO
      T A PIZZA.
2580  DATA  ITALY
2590  DATA  THEY ORDERED SPAGHETT
      I
2600  DATA  WHEN THEY LIVED IN P.
      E.I. DOLORES AND  BUTCH DR
      OVE A CHEVY.
2610  DATA  OWNED
2620  DATA  THEY LIVED IN P.E.I.
```

```
2630  DATA  BILLY AND DARLENE LIK
      E MICHAEL JACKSON  BUT SHE
      CAN'T STAND BOY GEORGE.
2640  DATA  SIT
2650  DATA  DARLENE LIKES BOY GEO
      RGE.
2660  DATA  MELANIE LIKES TO SWIM
       WITH LANCE SO SHE  ORDERED
      A SWIMMING POOL.
2670  DATA  BOAT
2680  DATA  MELANIE ORDERED A POO
      L.
2690  DATA  BUD AND FAITH DRINK K
      OOL-AID AND HE ALSO LIKES TH
      NG.
2700  DATA  SWALLOW
2710  DATA  BUD HATES THIS.
```

APPENDIX H: SPART LISTING, EXPERIMENT 3

```
10  REM  CREATED JAN 10,1984
20  REM  BY LUIS ZEGARRA
30  REM  REACTION TIME
40  REM   MODIFIED JULY 2ND, 1984
45  REM  PART ONE OF TEST
50  DIM BLOCK$(45,3),NUM(45),BN(4
    5)
60  DIM TIME(45,3),A11$(45,3)
70  HOME
80  PRINT "PLEASE ENTER THE FOLLO
    WING"
90  PRINT "IN ONE LINE (NO SPACES
    )"
100  PRINT : PRINT "INITIALS (2 L
    ETTERS)"
110  PRINT "AGE (2 DIGITS)"
120  PRINT "SEX (M OR F)"
121  PRINT "SUBJECT NUMBER"
122  PRINT "LIST NUMBER"
123  PRINT "LEFT OR RIGHT HANDED?
    "
130  INPUT CODE$
140  HOME : VTAB 15
150  HTAB 13
160  PRINT "PLEASE WAIT"
170  REM  READ SENTENCES INTO BLO
    CKS
180  FOR J = 1 TO 45
190  FOR I = 1 TO 3
200  READ BLOCK$(J,I)
210  NEXT I
215  READ L$: READ L$: READ L$
220  NEXT J
230  REM  RANDOMIZE SEQUENCE OF P
    RESENTATION
240  LET M = 0
250  FOR I = 1 TO 45: LET NUM(I) =
    0: NEXT I
260  LET N =  INT (45 *  RND (1))
     + 1
270  IF NUM(N) (  ) 0 THEN  GOTO
    260
280  LET NUM(N) = 1
290  LET M = M + 1
300  LET BN(M) = N
310  IF M (  ) 45 THEN  GOTO 260
320  REM  DISPLAY SENTENCES AND R
    ECORD RESPONSE TIMES.
330  HOME
340  FOR M = 1 TO 45
350  LET J = BN(M)
360  HOME
370  VTAB 15: HTAB 7
```

```
380  PRINT "..................
     ...."
390 X =  PEEK (49152)
400  IF X < 128 THEN  GOTO 390
410  GET B$
420  FOR I = 1 TO 3: HOME
430  VTAB 15
438  IF I = 3 THEN  FOR D = 1 TO
     750: NEXT D
439  IF I = 3 THEN  VTAB 20: HTAB
     8
440  IF I = 2 THEN  HTAB 16
441  IF I = 2 THEN  PRINT "+"
442  IF I = 2 THEN  FOR D = 1 TO
     300: NEXT D
443  IF I = 2 THEN  HOME : VTAB 1
     5: HTAB 16
450  PRINT BLOCK$(J,I)
460  NORMAL
470  VTAB 23
480  & CLRTIMER
490 X =  PEEK (49152)
500  IF X < 128 THEN  GOTO 490
510  & TIMERIN,(TV) = 0
520 Z =  PEEK (49168)
530 X = X - 128
540 A$ =  CHR$ (X)
550  REM  21 IS MEASURED PROCESSO
     R'S EXECUTION TIME
560 TIME(J,I) = (Q - 21)
570  IF A$ = "K" THEN  GOTO 630
571  IF A$ = "D" THEN  GOTO 610
580  IF A$ = "F" THEN  GOTO 630
581  IF A$ = "L" THEN  GOTO 610
590  LET A11$(J,I) = " "
600  GOTO 640
610  LET A11$(J,I) = "TRUE"
620  GOTO 640
630  LET A11$(J,I) = "FALSE"
640  IF TIME(J,I) > 10000 THEN  LET
     TIME(J,I) = 0 - 999
650  NEXT I
660  NEXT M
670  HOME
680  VTAB 13
690  HTAB 15
700  FLASH : PRINT "THANK YOU": NORMAL

740 D$ =  CHR$ (4)
745  ONERR  GOTO 940
750  PRINT D$;"OPEN"CODE$".P1,D2"

760  PRINT D$;"WRITE"CODE$".P1"
764  FOR J = 1 TO 45
765 M = J * 2 - 1
767  FOR I = 1 TO 3
770  PRINT M"."I;TIME(J,I);A11$(J
     ,I)
780  NEXT I
790  NEXT J
800  PRINT D$;"CLOSE"CODE$".P1"
810  POKE 49624,1
820 I$ =  CHR$ (9)
830  PR# 1
840  PRINT I$;"100N"
850  PRINT CODE$: PRINT : PRINT
```

```
860  FOR J = 1 TO 40
865 M = J * 2 - 1
870  FOR I = 1 TO 3
880  PRINT M"."I,TIME(J,I)"SECOND
     S",A11$(J,I)
890  NEXT I
900  PRINT : PRINT
910  NEXT J
920  PR# 0
930  GOTO 965
940 Y =  PEEK (222)
941  IF  NOT (Y = 4 OR Y = 8 OR Y
      = 9) THEN  GOTO 750
943  PRINT  CHR$ (7): HOME
945  VTAB 13: HTAB 5
948  IF Y = 8 THEN  PRINT "*** CL
     OSE DOOR OR INIT DISK ***"
950  IF Y = 4 THEN  PRINT "*** DI
     SK WRITE PROTECTED ***"
953  IF Y <  > 9 GOTO 961
955  PRINT D$;"DELETE"CODE$".P1"
958  PRINT "***DISK FULL - CHANGE
      DISK***"
961  PRINT : PRINT : PRINT :"PRES
     S RETURN TO CONTINUE"
963  GET Y$: PRINT
964  RESUME
965  END

]
```

APPENDIX 1:

Instructions, Experiment 3

First are you left- or right-handed (adjust template).

In this experiment you will read 90 sentences on the computer screen, and answer two true/falsequestions for each. Let me demonstrate with one of the 15 practice items. We begin with a row of dots and the MOVE instruction. Press the bar marked MOVE with your thumb to view the sentence. Read it once, and when you have understood it, hit "move" again right away. This brings up a cross for eye fixation, followed automatically by a probe word. With the appropriate finger, hit the T or F button to indicate whether or not that exact word was in the sentence you have just read. Finally, you will see a T/F question which I have inserted simply to ensure that you understand the sentences. Let's look at the practice items now...

APPENDIX J: Response Coincidence Analysis

Experiment 1

As suggested in the General Discussion of Experiment 1, an attempt was made to examine both subject strategy groups and item response clustering by a new statistical technique known as Response Coincidence Analysis (Baker and Derwing, 1982). Baker and Derwing have demonstrated the utility of this procedure for identifying stages in the acquisition of rules for plural inflections in English, and emphasize that working from similarities across items rather than grouping by age gives a much clearer picture of the organization of the linguistic data. They claim that an empirical and objective data analysis procedure must be used if we hope to identify groups of subjects who perform in similar ways with respect to a given set of items. In the case of pluralization rules, they were able to distinguish several "performance groups", composed of children of a variety of ages, thus resolving the "age-stage" confounding which makes the identification of developmental sequences so difficult in studies of language acquisition.

The technique was applied to the data of Experiment 1 in order to determine on the basis of similarities in responses to the "Crucial" items whether it would be possible to identify strategy groups - groups of subjects who differed either in the items which they rejected, or in terms of the number of items rejected.

Details of the procedure are given in the paper by Baker and Derwing. Results of the analysis are presented in Figures 1 - 6.

Figure 1 shows the four significantly different groups of subjects labelled Group 1 - 4. Unfortunately, no externally validated criteria (such as verbal I.Q., reading scores, vocabulary size, etc.) were available, but it would be of interest in later studies to attempt to explain subject groupings in terms of these kinds of individual differences in language abilities.

Figure 2 shows item clustering across all 30 subjects. There are only two significantly different groups (labelled Group 1 and Group 2) of items. These correspond to items either accepted as coreferential

(Group 2) or rejected as noncoreferential (Group 1). This interpretation is made on the basis of the appearance of a majority of the clearly acceptable control items in one group (items 61-75) and the clearly unacceptable controls in the other (items 45-60).

Figures 3 - 6 illustrate the four different classifications of the items by the four subject groups. In each case, only the main split into two subgroups is significant. This in itself is interesting, since it indicates that the target items were not treated as a separate type (e.g., "unusual" or "ambiguous" or "uninterpretable") by any of the four subject groups. Rather, we can characterize their item classification as a simple binary division between acceptable and unacceptable coreference, with the crucial items falling into one or the other category depending on the particular subject group in question.

For example, Group 1 subjects treated 14 of the 33 crucial items (2, 3, 4, 6, 7, 10, 12, 13, 17, 21, 28, 31, 42) in the same way as the clearly unacceptable items (45 - 60). Group 2 subjects treated 17 items like unacceptable ones (2, 3, 4, 6, 10, 13, 14, 17, 19, 22, 25, 28, 32, 33, 38, 40, 45), as did Group 3 (2, 3, 4, 5, 10, 13, 16, 17, 21, 22, 28, 29, 31, 33, 35, 36, 43). However, Group 4 treated only nine of the crucial items in the same way as the clearly noncoreferential items (2, 3, 4, 10, 13, 17, 19, 25, 26).

All four groups discriminated between the clearly acceptable and clearly unacceptable items: in each case, at least 14 out of 15 unacceptable control items were grouped together. The groups differed slightly in the number of misclassifications of the acceptable control items: Groups 1 and 2 classified only 1 clear YES with the NO items; Group 4 had two such misclassifications, and Group 3 had five.

Given this overview of the cluster data, we may focus on the interpretation of the qualitative differences between the subject groups. It would appear that the only real difference has to di with the number of target items that were classified as nonanaphoric. The appearance of only nine such items out of a possible 33 in the "unacceptable" cluster for Group 4 suggests that this is a very liberal group of subjects with respect to acceptabnce of coreference in unusual cases. Contrast this

with nearly double that number in Groups 2 and 3, and nearly 60% more in Group 1.

There is nonetheless evidence of consistency across groups in terms of the specific items that were rejected. Of the nine target items rejected by the most liberal group, six were also rejected by all three of the other groups; one by two other groups; and two by one other group. This suggests real differences in the subjects ability to accept coreference in particular.

Differences across groups are more difficult to characterize than their similarities, since the items were not sampled and replicated with this sort of analysis in mind. However, the approach does seem promising, and will be considered in future work in this area.

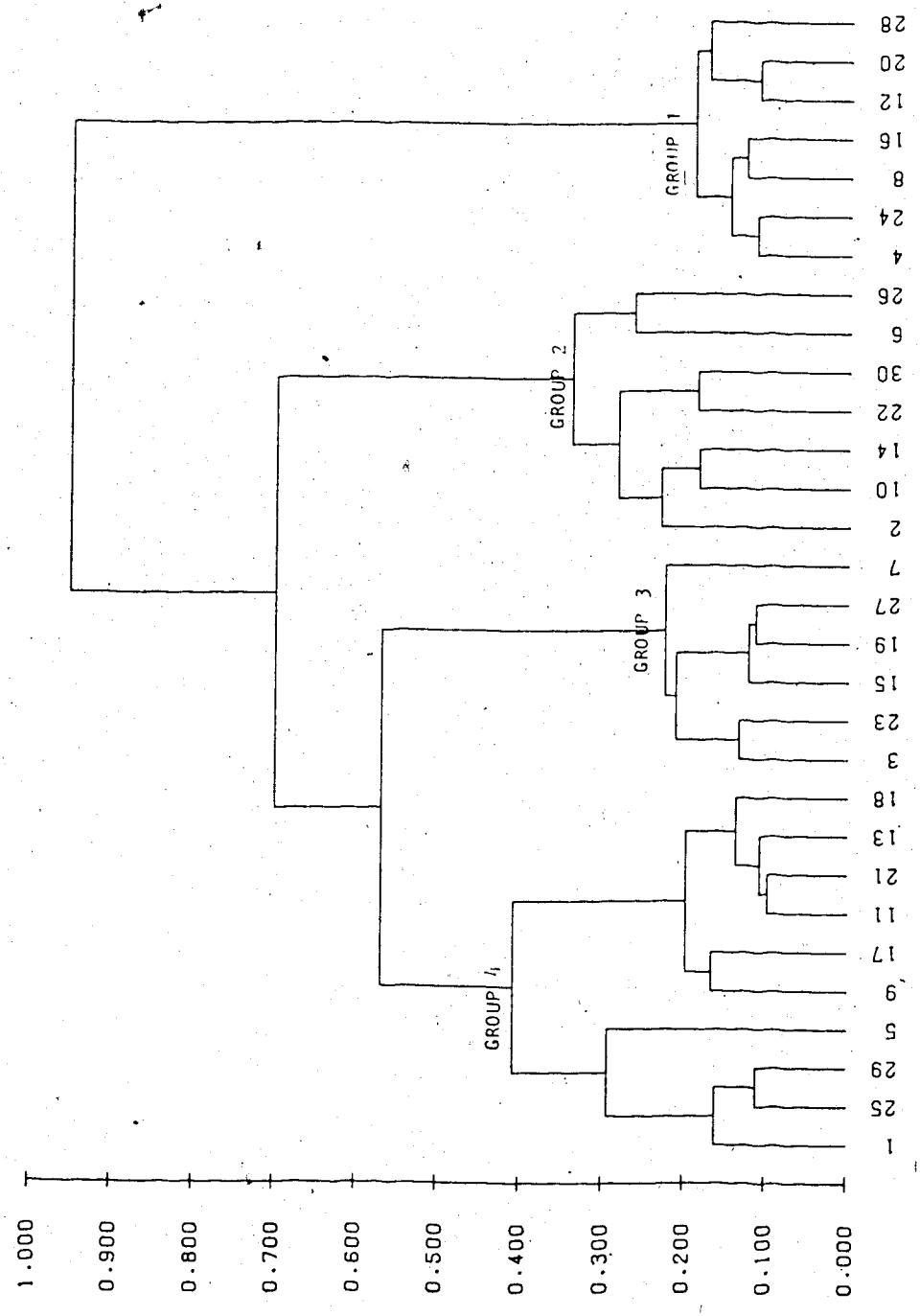


Figure 1: Subject groups, Experiment 1.

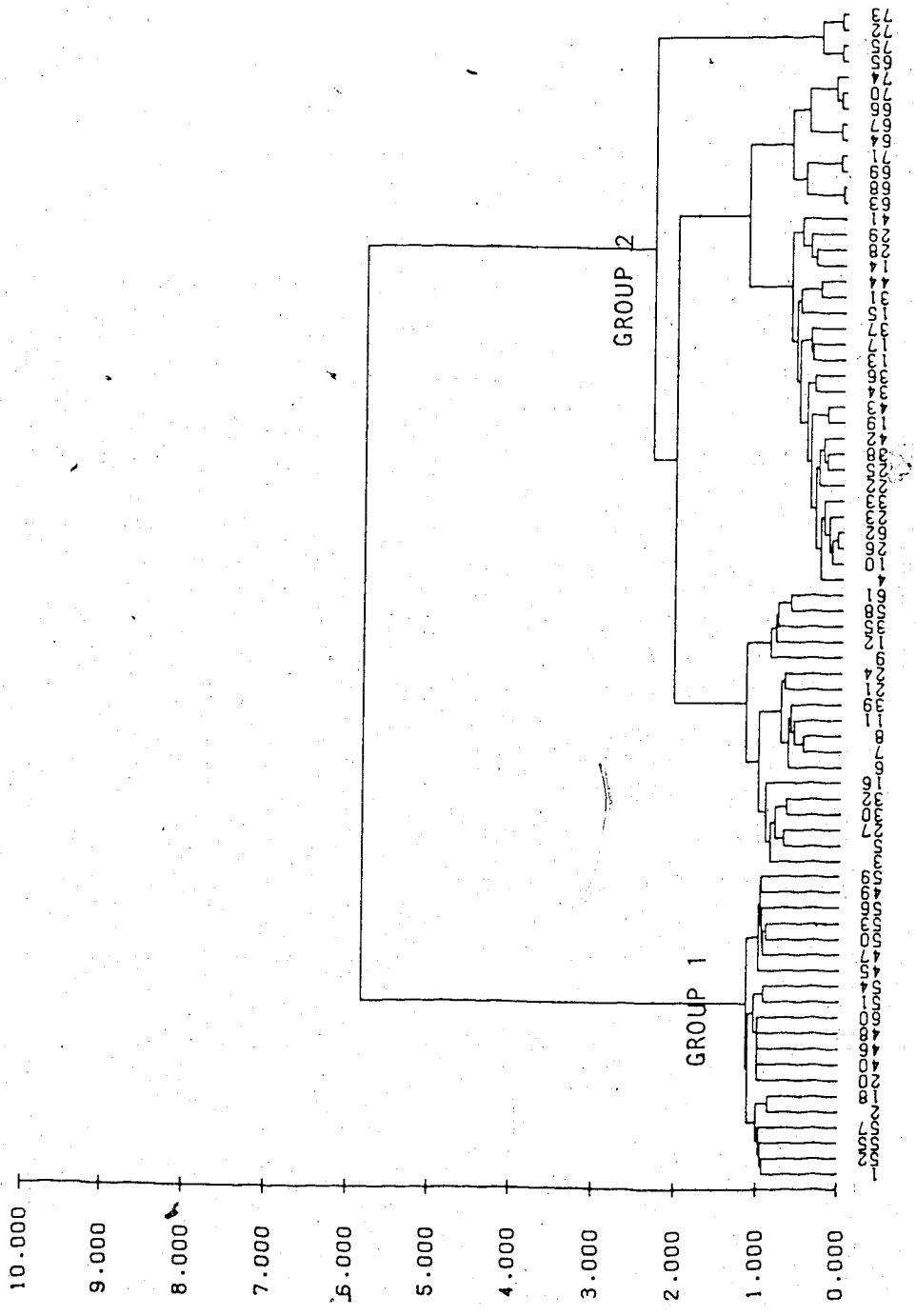


Figure 2: Item clusters across subject groups, Experiment 1.

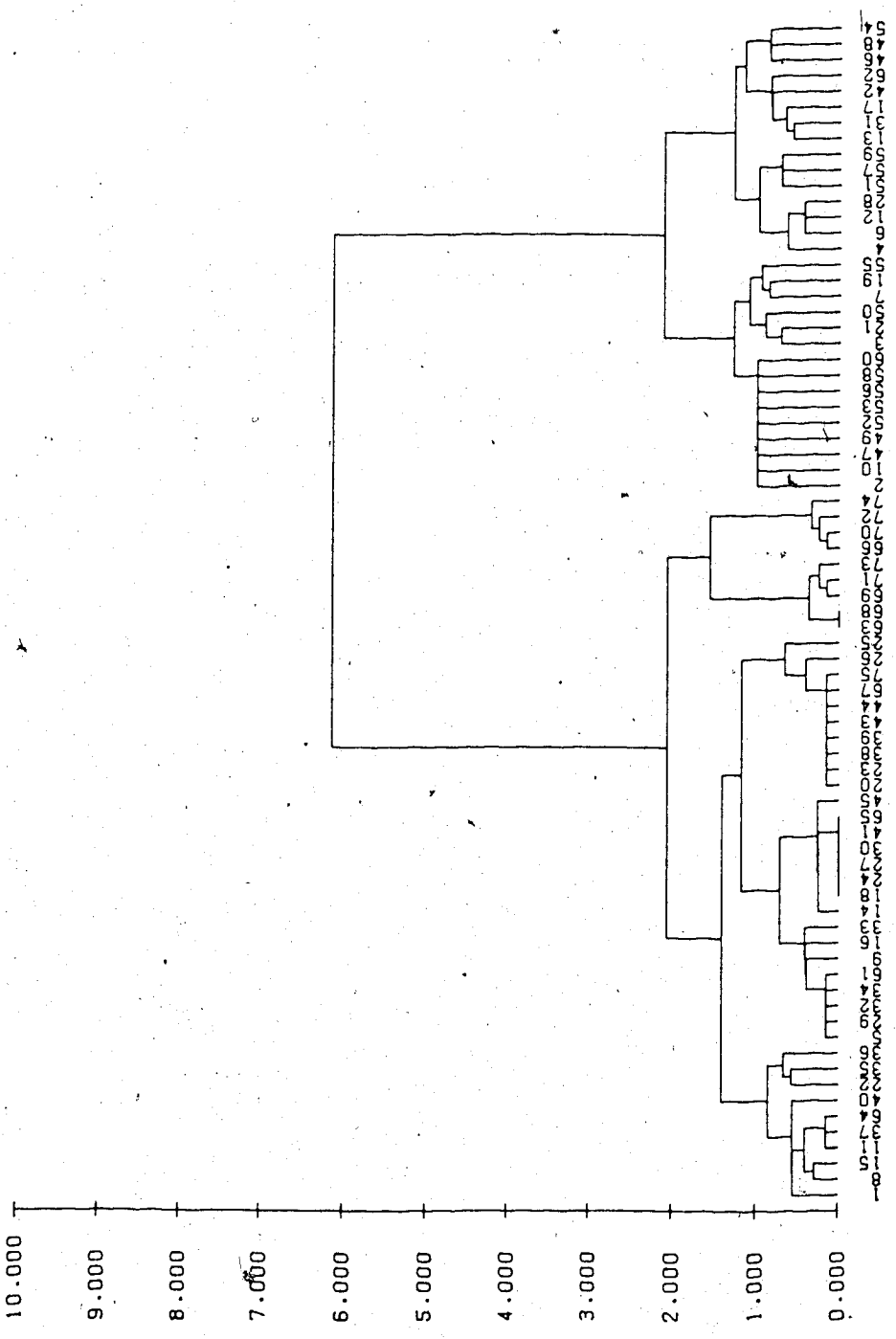


Figure 3: Item clusters for Group 1 subjects.

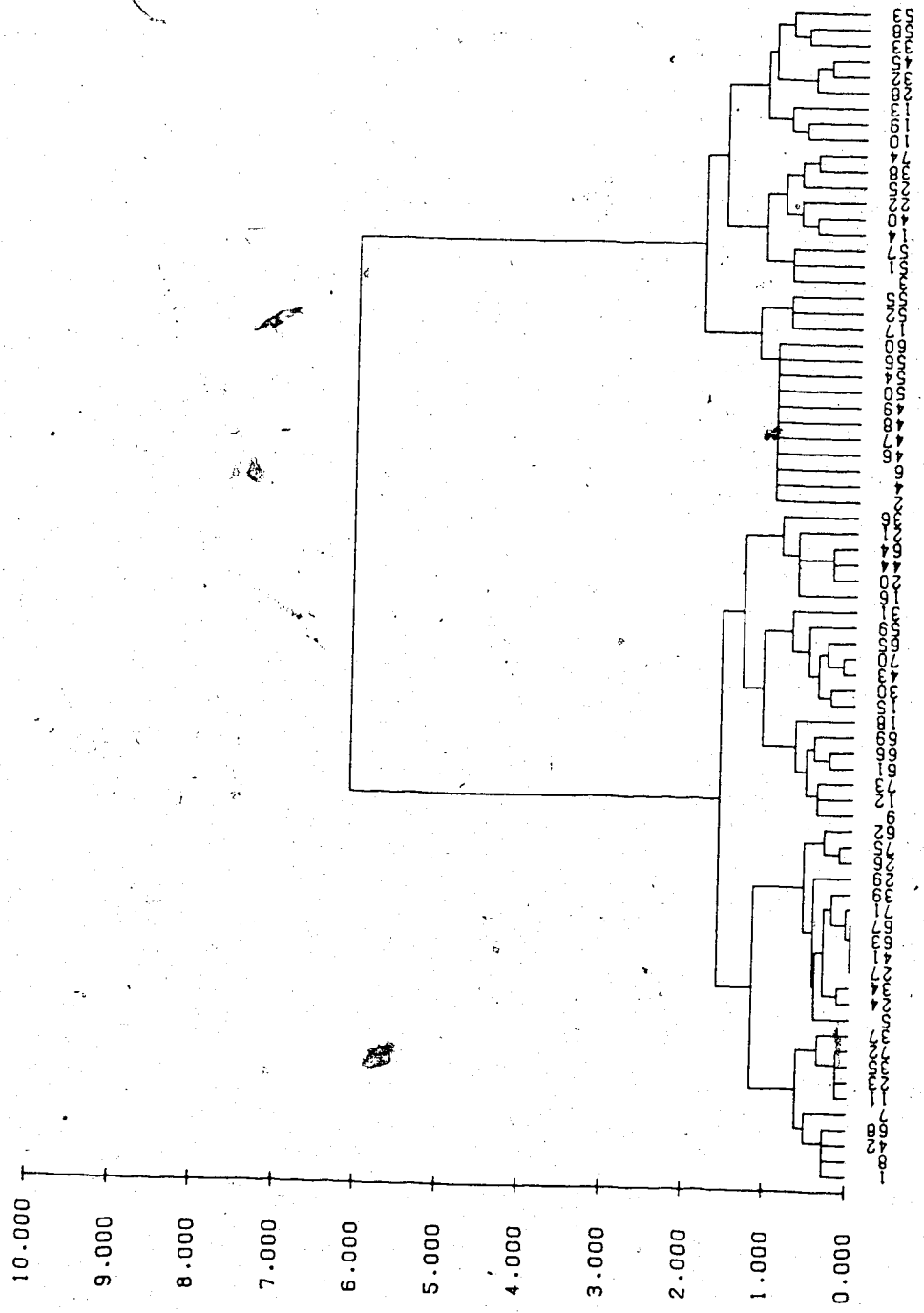


Figure 4: Item clusters for Group 2 subjects

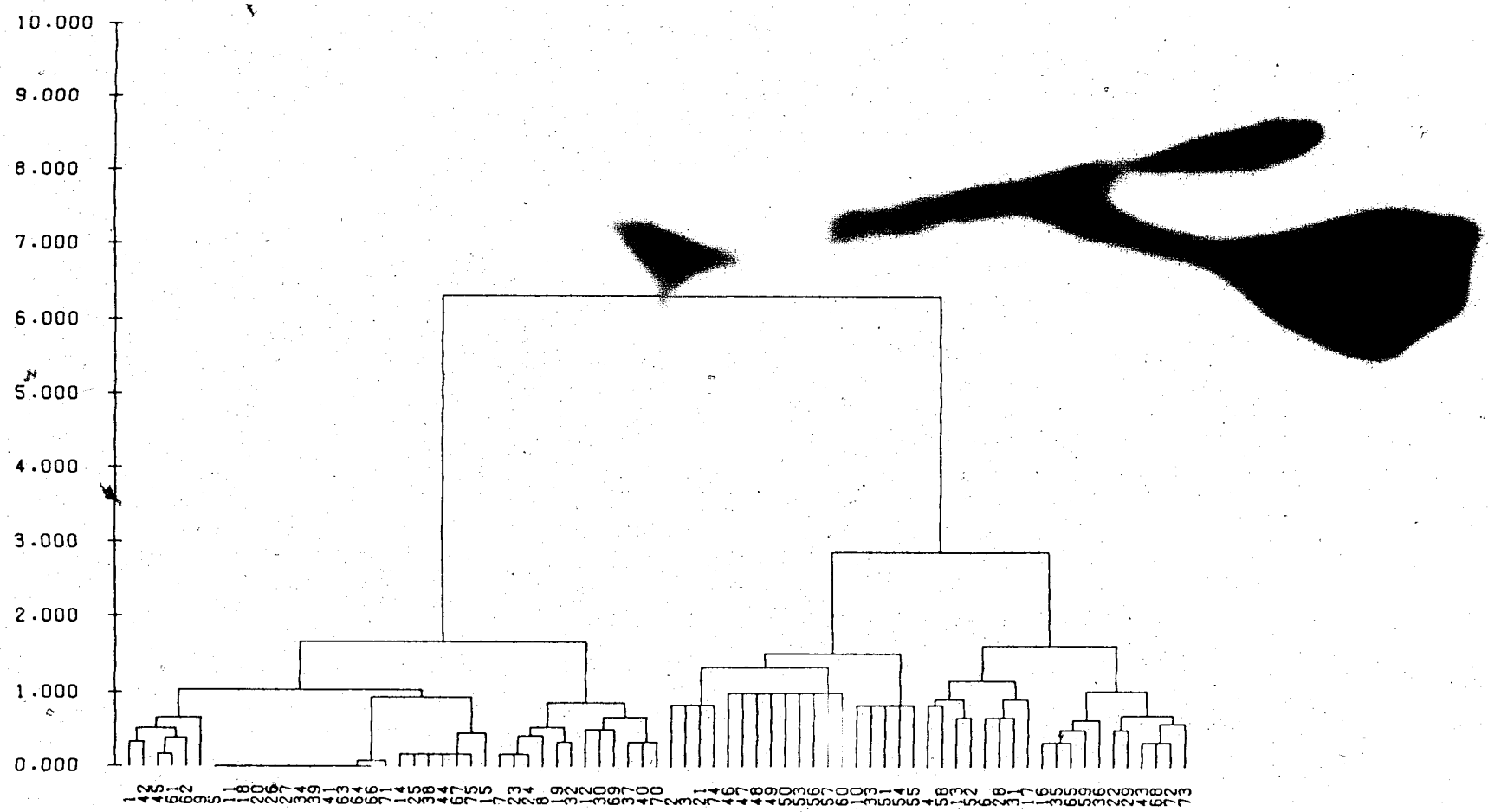


Figure 5: Item clusters for Group 3 subjects.

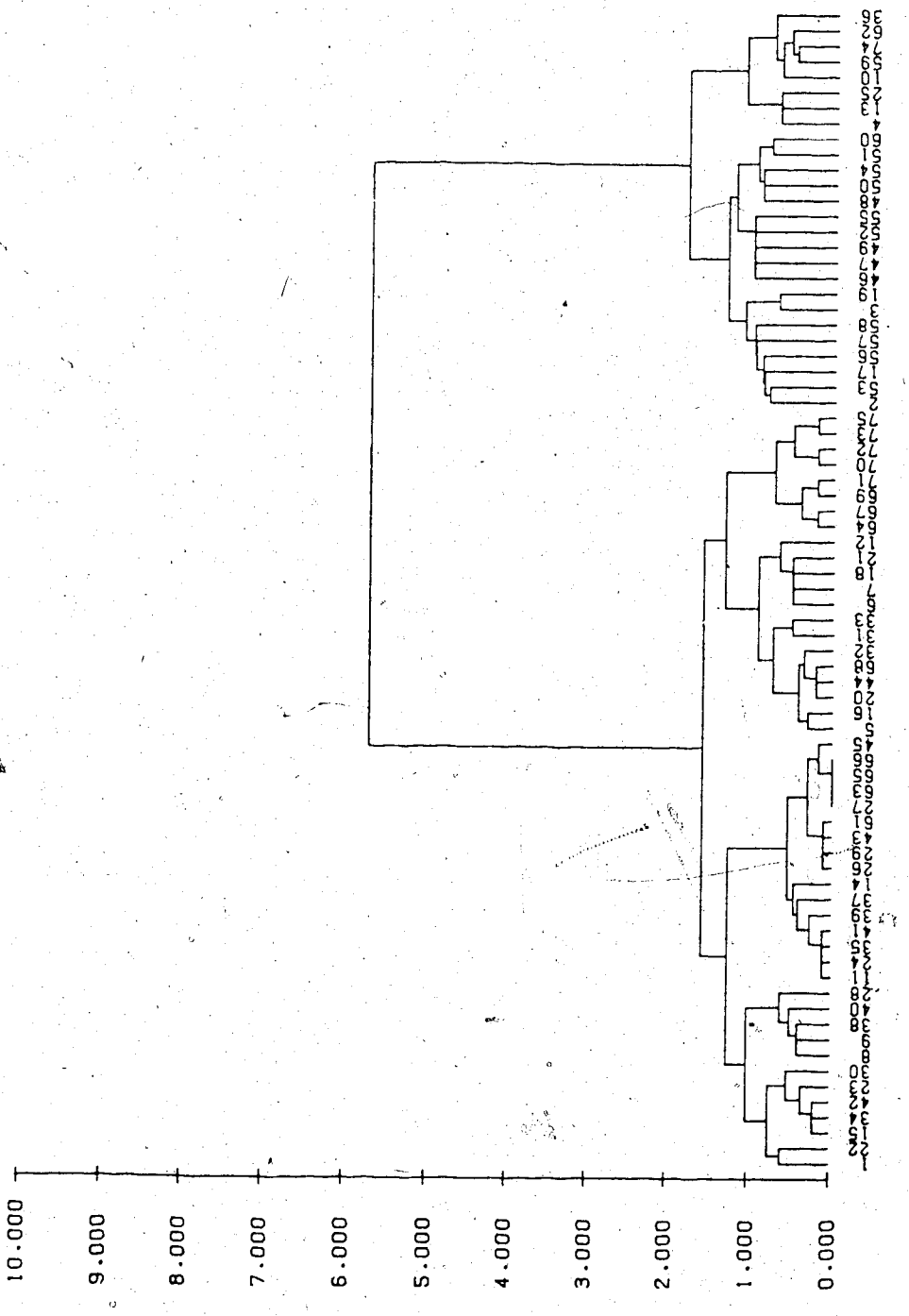


Figure 6: Item clusters for Group 4 subjects.