The Value of an Annotated Corpus in The Investigation of Anaphoric Pronouns, with Particular Reference to Backwards Anaphora in English.

Izumi Tanaka

Abstract

This thesis investigates English personal pronoun reference in particular focusing on **cataphora (backwards anaphora)**, using **the Anaphoric Treebank (AT)**, which is a written English corpus with discourse annotation, and other corpus data. The analysis of corpus data reveals certain coreferential cataphora patterns (in particular in the initial adverbial or initial direct speech constructions). On the basis of the corpus data, the claims on cataphora made by generative approaches and cognitive discourse theories are tested.

The points which became clear in testing generative approaches are:

(1) The result of an informant test indicates the inadequacy of narrowly restricting data to invented examples.

(2) The lack of understanding of the scope of the application of the theory can be observed in Reinhart (1984) and Binding theories.

(3) A sentential-level approach to investigate pronoun references is inadequate.

The points which became clear in testing cognitive discourse theories are:

(i) First mention cataphora and Overriding-type of cataphora clearly show that there is such a phenomenon that can be called cataphora, contrary to the sceptical view on cataphora.

(ii) These cataphora data indicate that a pronoun reference can be made not only for an already-mentioned entity but also for a new discourse entity.

Also the analysis of the borderline cases (anaphora/cataphora) reveals that in certain conditions, the reference-direction (anaphora/cataphora) judgement tends to involve a triple-choice (anaphoric / cataphoric /indeterminate), or it can involve a cline between the anaphoric pole and the cataphoric pole. This indicates that the

reference-direction judgement of a pronoun (anaphora/cataphora) tends to reflect the way in which a reader perceive focus of attention (highlight) when the reader completes to resolve the pronoun.

In order to account for the cataphora data, a suggestion is made that it is necessary, from readers' perspective, to assume that a pronoun creates a temporal information gathering point in some kind of the reader's short term memory.

Dedication

This thesis is dedicated to

Tomoji Tanaka,

Mine Tanaka,

and

the founder of my alma mater,

Dr. Daisaku Ikeda.

Acknowledgements

In addition to the three people to whom this thesis is dedicated, there are so many people to whom I must express gratitude for the completion of this thesis, and unfortunately I could not include all of their names here.

First of all, I would like to thank my supervisor Professor. Geoffrey, N. Leech for his excellent supervision, patience and help during the course of writing this thesis. Without his invaluable comments and suggestions, I would never have been able to complete this research.

Thanks are also due to another supervisor, Dr. Anthony M. McEnery for his support in particular in the early stage of this research.

I must also express my gratitude to Professor. Anna Siewierska and Professor. Ruslan Mitkov for their invaluable comments, which enable me to improve this thesis significantly.

I must express my gratitude to Nicholas Smith for his great help in the last stage of writing up the thesis as well as giving me useful advice in making questionnaire forms. Also I would like to express my gratitude to Banti Zelalem for his invaluable comments on my draft.

I must thank the members of the research team who create the Anaphoric Treebank, which is the main data of this research. Thanks in particular go to Steve Fligelstone for providing me the invaluable documents about the coding and compilation of the Anaphoric Treebank.

I would like to express my gratitude to Michael Oakes for his support in accommodation, Marjorie Wood and Dorothy Barber for the administrative support, Reiko Ikeo and David Lee for their comments on the final draft, and for all the native speakers of English who kindly offered their intuition for my informant test.

I would like to express my sincere gratitude to Professor. Akira Yanabu, whose works have been always the source of inspiration for me, and Peter Davis for his encouragement and friendship since I came to Lancaster.

Table of Contents

Conventions and Abbreviations

(i) Quoted words or phrases in the narrative are typed in italics:

   e.g. . link between the 3PP *it* and the NP *the base salary* after *said of* ...

(ii) A relatively long citation is typed in italics and enclosed by double quotes:

   e.g. the 3PP *he* as being coreferential with the clause subject ***"The vice president of the West German Olympic Committee"***.

(iii) A discourse entity or the referent of a linguistic expression is enclosed by single quotes:

   e.g. 'Chris Evert Lloyd' is referred to with the expression *Chris Evert Lloyd* in the sentence ...

(iv) The coreferential NPs are typed either in bold or in italics or underlined. In the following example, there are three coreferential pairs: The NP **the president** and the 3PP **his**, The NP *the Iranian leader* and the 3PP *his*, and the NP the spokesman and the 3PP he:

> [A031 44] Pressed about the possibility of some contact between **the president** or **his** representatives and *the Iranian leader* or *his* representatives, the spokesman said he assumed "there could have been something, but this is something I don't know about."

(v) Instead of using rather confusing pair of terms addresser/addressee and intra-sentential/inter-sentential, I will use **addresser/receiver** and **intra-sentential/cross–sentential** respectively.

(vi) Each example sentence of the Anaphoric Treebank is preceded by the sentence identification enclosed with square brackets:

   e.g. **[A044 71]** "I'm really keen to ...

(vii) The bold italic *i* or *j* placed at the end of a word is a co-index (coreferential) marker.

       e.g.   They*i* saw a snake*j* near them*i*.

Table of Abbreviations:

(the) AT	The Anaphoric Treebank
CGEL	Comprehensive Grammar of English Language (Quirk Et Al, 1985)
NP	Noun Phrase
VP	Verb Phrase
PP	Prepositional Phrase
1PP	First Person Pronoun
1PSP	First Person Singular Pronoun
2PP	Second Person Pronoun
3PP	Third Person Pronoun
3PSP	Third Person Singular Pronoun
BT	Binding Theory
MC	Main Clause
MR	Mental Representation
initial-ADV	Initial Adverbial (an adverbial occurring at the initial position / preposed adverbial)
initial-DS	Initial Direct Speech (a piece of direct speech occurring at the initial position / preposed direct speech)
initial pronoun	A pronoun in the initial element (initial-ADV or initial-DS)

Chapter 1: Introduction: Background of the Thesis

1-1. Aims of This Thesis

This thesis presents a corpus-based investigation of English personal pronoun reference, in particular focusing on cataphora (backwards anaphora). This thesis has two main objectives:

Firstly, this thesis attempts to investigate the little studied phenomenon of cataphora (backwards anaphora) with corpus data. Although considerable work has been done on anaphoric pronoun reference, cataphoric use of pronouns has been little studied, in fact almost neglected (Kamp and Reyle 1993, Green et al 1994), except for backwards anaphora studies in the generative framework, which concentrate mainly on invented single sentences as their data. Hardly any of the generative approaches present and discuss naturally occurring discourse data. Also there is a prevailing scepticism towards the existence of cataphora (backwards anaphora) itself in the generative approach (Kuno 1972, 1975) as well as among other linguists (Delisle 1973, Bolinger 1977, Cornish 1996). As one of the main reasons for the scepticism or neglect is the lack of appropriate empirical discourse data, this thesis attempts to give an account of cataphora based in particular on a corpus with discourse annotation

(the Anaphoric Treebank: described in Chapter 3).

Secondly, by retrieving regular patterns of cataphora from the AT (Chapter 4), and by examining the claims about cataphora made by generative syntactic approaches (Chapter 5) and cognitive discourse theories (Chapter 6) against the examples in the AT, this thesis aims to demonstrate the usefulness of a corpus with discourse annotation such as the AT in the study of pronoun reference.

What is unfortunate for the empirical study of languages is that due to the dominance of the generative approaches, empirical authentic data have been largely excluded from the object of inquiry. This thesis in particular emphasises the use of corpus as a test bed, and examines the empirical validity of the various claims made about cataphora (backwards anaphora).

In the next section (1-2), I will discuss the relevant methodological aspects of corpus use, and an ideological aspect of the neglect of corpus use. In 1-3, I will briefly overview the issues related to cataphora (backwards anaphora), followed by the plan of this thesis in 1-4.

1-2. A Discussion of Corpus-Based Approach

1-2-1. A Discussion of Methodological Aspects of Corpus Use

There are different kinds of corpus created for different purposes: a corpus of news reportage, a corpus of conversation, a corpus of a particular writer etc. Each of them serves to provide empirical data for different kinds of inquiries.

The genre of the corpus used in this thesis is mainly news reportage, which is a kind of published informative texts intended to convey factual information. In the discussion of corpus in this section, I will mainly talk about a corpus of collection of published informative texts, such as newspapers or scientific papers.

For a corpus of published material, we can assume that the texts have gone through a revising and editing process before publication, which minimises the inclusion of what are called 'performance phenomena', such as slips of the pen/tongue or spelling mistakes. We can also assume writers' and editors' judgement to be that their texts are presented as sufficiently communicative and acceptable to the intended readers (though their judgements are not always adequate).

What we can directly observe in a corpus of published texts is not the process but the product of people's writing to communicate with others.[1] A corpus can provide empirical data for inquiries focusing on the product of writing, such as: how people used words and phrases, or constructed sentences in their writings; how people chose referring expressions (proper names, definite NPs, or pronouns), etc. Even for generative approaches, such as Carden's position discussed below, a corpus can

---

[1] See Leech (1992:108) for the distinction between 'product' and 'process' of 'performance'.

provide empirical evidence as to how people 'generated' sentences.

> ... after all, we want our grammar to account for what people actually say and
> write. (Carden, 1982)

For those inquiries for which a corpus provides empirical data, the methodological status of the corpus is twofold: one is the corpus as a **test-bed**, the other is the corpus as a **source of knowledge**. These two aspects need to be distinguished clearly.

As a **test-bed**, the status of a corpus is rather absolute in the sense that when empirical predictions made by a theory do not accord with the data in the corpus, it is not the data but the theory that has to be rejected, discarded or modified. Empirical theories, in this respect, have to be falsifiable.

In a corpus, a linguist may sometimes find some elements which s/he judges to be irrelevant to his/her inquiry, such as a particular writer's idiosyncratic usage or stylistic deviation. Yet s/he can never know whether other people also judge the elements as idiosyncratic or stylistically deviant. What the linguist can do, then, is to state explicitly his/her own judgement, referring to the elements in the data that s/he judges as idiosyncratic use of stylistic deviation. No empirical scientists or empirical theories can afford to discard or neglect empirical data. What is allowed for an empirical scientist is not to discard the data but to state clearly his/her own subjective judgement on the data.

The significance of the recent development of computer corpora, as illustrated by Leech below, is largely concerned with the second status of corpora, namely the corpus as a **source of knowledge**.

> ... the computer's ability to search, retrieve, sort, and calculate the contents of vast corpora of text ... at an immense speed, gives us the ability to comprehend, and to account for, the contents of such corpora in a way which was not dreamed of in the pre-computational era of corpus linguistics. ... they provide the basis on which, for example, complex mathematical models of linguistic behaviour can be built.
> (Leech, 1992:106)

The significance of computer corpora should be underlined also in terms of the first status of corpus (as a test-bed). Computer corpora considerably contribute to increase the sharability of data among researchers as well as to facilitate the creation of the record of analysis (corpus annotation), which improve the conditions of the testability of theories significantly.

Other than a corpus, linguistic models or knowledge can be obtained from a linguist's own intuition, academic publications, informant test results etc. In expressing his "denial of authority to the sources of knowledge", Popper says:

> .. every 'source' - tradition, reason, imagination, observation, or what not - is admissible and may be used, *but none has any authority.* ... every source is welcome, but no statement is immune from criticism, whatever its 'source' may be. (Popper, 1966: 378 -- emphasis by Popper)

As Popper noted, none of the sources of knowledge can claim its superiority over others. A linguistic model or knowledge cannot claim superiority over others because of its source of information. Intuition-based models cannot claim superiority against corpus-based models, and vice versa. What matters is not the source from which the model of knowledge is obtained, but whether a model or knowledge can made a prediction about what is likely to be observed in empirical phenomena. Any models or knowledge have to be tested against a corpus (or some other test-bed) if they claim

to contain empirical scientific statements at all, as Popper states:

> In so far as scientific statements refer to the world of experience, they must be refutable; and, in so far as they are irrefutable, they do not refer to the world of experience. (Popper, 1966: 13)

Although a corpus of publications (an unannotated 'raw' corpus) enables us to observe the products of people's writing directly, we cannot directly observe the process and the results of how people interpret the products (namely texts in a corpus). Here we face the classical philosophical problem of **other mind**, i.e. no one can gain access to or directly observe anybody else's mind or intuition. An arm-chair linguist can only have access to his/her own intuition. S/he can never be able to access other people's intuitions or grammatical judgements.

If a linguist tries to investigate the issues of text interpretation solely on his/her own, it may be enough for him/her to invoke his/her own intuition. Yet in order to investigate the issues of interpretation more objectively, with other people, what we need is the record of each person's interpretation of texts, based on which we can pursue our inquiries with the aid of others.

In practice, we may need "another source of empirical data for linguistics - the use of informant tests" (Leech 1992: 116), or corpus annotation, i.e. "the practice of adding interpretative, linguistic information to an electronic corpus of spoken and/or written language data" (Garside, Leech and McEnery 1997: 2). The discourse annotation of the AT, where the analytic judgements of native speakers are recorded through the annotations, is one type of record of the interpretation of texts.

Because of its psychological nature, careful data assessment may be needed in using the record of interpretation. Yet no matter how much difficulty may be

involved in obtaining the record of interpretation, a linguist certainly needs it if s/he

tries to investigate interpretative issues not solely her/himself, but with other people.

An annotated corpus such as the AT is particularly valuable in containing a record of

both the productions of writers and the judgements of interpreters of the same data.

One of the main aims of this thesis is to demonstrate the usefulness of a corpus, in

particular a corpus with discourse annotation, in the study of pronoun reference. This

thesis emphasis the use of a corpus as a test-bed and tests various claims made on

cataphora from different theoretical or analytic point of view.[2]

---

[2] In the discussions of generative approaches (chapter 5), I will invoke not only
corpus examples but also invented examples judged as acceptable by native speakers.
This is mainly because it is usually difficult to obtain the corpus examples which can
be used to test the claims made by generative linguists.

     I do not take the extreme position which excludes any invented sentences from
the empirical data of linguistics. since. after all. any corpus sentences are also
invented by humans. In the view of linguists' invented examples, my approach differs
from many of the generative approaches in the following points:
(i) Whereas many generative linguists tend to present isolated single sentences as
their data. I emphasise that a sentence example should be presented with its co-
occurring text, as much as possible.
(ii) Many generative linguists tend to present their examples to be acceptable or not
on the basis of their own introspection. My emphasis. on the other hand. is that
linguists' invented sentences need to be presented with the qualification that how
many agreements and disagreements are involved in the acceptability of their
sentences. (see Brown and Yule. 1983:21 --- quoted in the next section) My invented
examples in chapter 5 are presented with the acceptability judgement made by 15
native speakers of English in Lancaster University.
     Strictly speaking. a corpus example also needs to be presented with the voices
of acceptability judgement. In this thesis. I will present the corpus examples
(published texts) on the basis of the acceptability judgement of myself and of the
analysts of the Anaphoric Treebank. as well as assuming that the writers and editors
judged their texts to be sufficiently acceptable to the pubic.

1-2-2. A source of non-corpus based approach

In referring to the appearance of the term 'corpus linguistics', Leech notes:

> I am not aware that it made its appearance in the corpus linguistics heyday of the 1950s - simply because, for those who espoused this approach, corpus linguistics was simply "linguistics" - to them, no other linguistics deserved the name. (Leech, 1992:105)

For those who engaged in linguistics research in 1950s, it would be simply impossible to imagine any discipline of linguistics, if it claims to be a discipline of empirical science at all, which can afford to neglect empirical data. Such an inconceivable move, the generative approach, was introduced to linguistics by Chomsky in the late 50s. Since then, the empirical investigation of language has suffered from the rise of the generative approach.

One of the main causes of neglecting empirical data in the generative framework is the distinction between **competence** and **performance** made by Chomsky. There are extensive discussions made about the dichotomy of competence and performance in terms of Linguistic methodology, and I cannot go into details of them here. In this section, I will focus on an ideological aspect underlying Chomsky's **competence/performance** dichotomy, i.e. the influence of **Platonic idealism**, which is the main factor underlying Chomsky's neglect of empirical data.

The influence of Platonic idealism on Chomsky, particularly with regards to the competence/performance dichotomy, is pointed out by linguists such as Sampson:

The competence/performance distinction implies that the messy complexity of NL as it is actually encountered on the page or in speech masks a far more orderly, elegant Platonic ideal language lurking below the surface --- and it implies furthermore that the proper object of a linguist's attention is the elegant ideal competence rather than the real but relatively intractable performance.
(Sampson, 1987:27)

A concise definition of Platonic idealism can be found in the following comment by

Popper on the labour theory of value presented by Karl Marx:

The whole idea -- which was not Marx's invention -- that there is something behind the prices, an objective or real or true value of which prices are only a 'form of appearance' show clearly enough the influence of Platonic idealism with its distinction between a hidden essential or true reality and an accidental or delusive appearance.
(Popper, 1966:177)

We can see that Chomsky's competence/performance distinction corresponds to the

distinction between 'hidden essential or true reality' (corresponding to 'competence') and

'accidental or delusive appearance' (corresponding to 'performance') in Platonic idealism.

From the point of view of empirical inquiry, the problem of Platonic idealism is

that it tends to recognise empirical phenomena to be 'delusive' or 'accidental' and exclude

it from the object of inquiry. The exclusion of empirical phenomena from the object of

inquiry is clearly expressed in the following passage by Chomsky:

Observed use of language or hypothesised dispositions to respond, habits, and so on, may provide evidence as to the nature of this mental reality, but surely cannot constitute the actual subject matter of linguistics, if this is to be serious discipline.
(Chomsky, 1965:4)

Such a tendency can be observed not only in Chomsky but also, as pointed out by Popper

above, in Karl Marx's statement in *Capital*:

> If prices actually differ from values, we must first reduce the former to the latter, i.e. disregard this situation as an accidental one in order to observe the phenomenon of the formation of capital on the basis of the exchange of commodities in its purity, and to prevent our observations from being interfered with by disturbing incidental circumstances.
> (Marx, 1976:269)

Notice that Marx talks about the 'reduction' of 'accidental' empirical phenomena (prices) to the 'pure' reality ('value'). This is a typical sign of the influence of Platonic Idealism, that can also be observed in Chomsky:

> Actual investigation of language necessarily deals with performance, with what someone does under specific circumstances. We often attempt to devise modes of inquiry that will reduce to a minimum factors that appear irrelevant to intrinsic competence, so that the data of performance will bear directly on competence, as the object of our inquiry.
> (Chomsky, 1980:25)

Notice that, in a similar way to Marx. Chomsky talks about 'reduction' of empirical data to "intrinsic" entity ('competence'), which is the object of his inquiry. For Chomsky, the object of observation is not the naturally occurring empirical data which exhibit how and what people actually wrote and talked in their communication, but either the sentences invented by a linguist or empirical data 'reduced' by a linguist. For Chomsky, empirical data cannot be the object of a linguist's observation without being 'reduced' by his/her private introspection, which alone can gain access to the object of inquiry ('competence'). An actual consequence of this limiting of the object of observation is pointed out by Brown and Yule's comment:

> The essential problem in an extreme version of the constructed-sentence approach occurs when the resulting sentences are tested only against the linguist's introspection. This can (and occasionally did) lead to a situation in which a linguist claims that the 'data' he is using illustrates acceptable linguistic strings because he says it does, as a result of personal introspection, and regardless of how many voices arise in disagreement.
> (Brown and Yule, 1983:21)

A problem of an extreme version of the constructed-sentence approach will be looked at in chapter 5, with the result of an informant test.

Empirical sciences fundamentally differ from Platonic idealism in the view of empirical phenomena. An empirical scientist does not regard empirical phenomena as simply 'disturbing', 'delusive', or 'accidental'. The central belief underlying empirical scientific investigation is that no matter how 'chaotic' empirical phenomena appear to be, they are in fact governed by laws which careful examination of the phenomena can reveal. It is the observable regularities that are the main object of inquiry in empirical sciences.

1-3. Issues Related to Cataphora (Backwards Anaphora)

In the majority of cases, pronoun reference is anaphoric. (cf. 3-8) The referent of a pronoun is usually assumed to be already recoverable or accessible in some way, either in the immediate context or preceding text or reader's memory (short term or longer term). In this respect, anaphora is the norm for pronoun reference. Cataphora, i.e. where a pronoun appears before its coreferential item, is not as common as anaphora. Cataphora is usually said to be "associated with formal written English"(CGEL: 351). Although cataphora has not been paid as much attention as anaphora by linguists, several syntactic approaches were made to cataphora (backwards anaphora) in the generative framework (e.g. Langacker 1969, Ross 1969, Reinhart 1984, Carden 1986) under their terms of 'pronominalisation' or 'backward pronominalisation'. Those syntactic approaches try to capture the syntactic conditions or syntactic 'domain' within which the coreference relations between NPs are constrained. In chapter 5, I will empirically examine and discuss C-command coreferential constraints (Reinhart, 1984) in 5-1, Backward pronominalisation rules (Langacker 1969, Carden 1986) in 5-2, and Binding theory (Chomsky, 1981, 1985, 1986) in 5-3.

There exists, on the other hand, a scepticism about these syntactic approaches to cataphora and even about the existence cataphora itself either within the generative framework (e.g. Kuno 1972,1975) or outside of the framework (e.g. Bolinger 1977, Cornish 1996. Stockwell, 1995). Kuno and Bolinger question whether those cataphora 'data' presented and discussed in the syntactic approach are really cataphoric. In my

view, there are good reasons for this scepticism. Firstly, for those who are outside the generative framework (e.g. Bolinger 1977), it is hard to accept the notions of 'pronominalisation' and 'backward pronominalisation' which necessarily assume the theoretical postulations of a generative framework, such as 'deep/surface structure', 'base generated' or 'movement'. Secondly, most of the accounts of cataphora made within generative framework make use of invented, isolated sentences as their 'data' of cataphora. Hardly any of them confront corpus data, presumably due to their methodological position thereby empirical data tend to be neglected as 'performance'. If we observe, however, a single sentence with its co-occurring context, the seemingly cataphoric case sometimes can also be analysed as anaphoric.

Consider:

[A022 17] When **he** retired , **DeLury** was succeeded by the union 's secretary-treasurer, Edward Ostrowski, 56 , who said of his death: "We have lost a major figure in our lives".

The male pronoun *he* in the initial adverbial clause can be analysed as cataphoric, being coreferential with the following main subject *DeLury*. If this isolated sentence is observed in its concurring context, however,

[A022 16] Over the years, **DeLury** gained a reputation on both sides of the bargaining table as an effective labor leader and an innovative negotiator.
[A022 17] When **he** retired , **DeLury** was succeeded by the union's secretary-treasurer , Edward Ostrowski, 56 , who said of his death: "We have lost a major figure in our lives"

we can see that the male pronoun *he* in [A022 17] can also be analysed as anaphoric,

being coreferential with the main subject *DeLury* in the previous sentence [A022 16].[3]

In its strong form (e.g. Kuno 1972), the scepticism towards cataphora asserts that there is no 'genuine' cataphora (backward anaphora) in its literal sense; all the seemingly cataphoric pronouns must have their coreferential expressions somewhere in the preceding context. This strong position is, however, refuted forcefully by Carden (1982) who provides rather conclusive empirical evidence against the extreme scepticism, using his own over 800 examples. His examples include those cases where the pronoun is "the first mention of its referent in the discourse" (Carden, ibid.). This kind of cataphora, i.e. a pronoun referring to first-mentioned entity in the discourse (henceforth 'First-mention' cataphora), can also be found in the Anaphoric Treebank (AT):

[A033 32] Amid demands for **his** resignation, **Democratic state Rep. Russell J. Reynolds** apologized Friday for using a racial slur in a response to a survey.

Since this sentence appears at the beginning of a text, there is no way to interpret the 3PP *his* as anaphoric. Those corpus data of cataphora are presented in chapter 4. As we will discuss in chapter 6, the corpus data of cataphora such as the first-mentioned cataphora seems to have certain theoretical impact upon those cognitive discourse theories which assume that pronouns usually refer to highly recoverable or accessible entities.

---

[3] This 3PP in [A022 17] is judged as anaphoric by the analyst of the Anaphoric Treebank (AT), encoded *<REF=1 he*.

Scepticism towards cataphora is addressed also from the point of view of on-line (sequential linear) text analysis. Since readers usually process a text sequentially or mono-directionally (from the beginning to the end of a text), the notion of reference-direction (anaphora/cataphora) appears to be implausible if it is taken to mean in a formal textual way, such as 'a pronoun refers back/forward to the antecedent NP'. I will discuss how the notion of reference-direction (anaphora/cataphora) can be accounted for in terms of on-line text analysis in chapter 7.

1-4. What follows in This Thesis.

In chapter 2, as a foundation of the discussions in the rest of the thesis, I will overview several factors and issues relevant to the interpretation of a personal pronoun in news reportage, which is the genre of the data used in this thesis. Topics in the chapter include: deictic dependency of interpretation of personal pronouns; mutual exclusive relations of 'person' in the interpretation of personal pronouns in direct speech; a third person pronoun (3PP) as a signal for highlight (focus) maintenance; conceptual issues related to re-highlighting.

Chapter 3 describes the Anaphoric Treebank (the AT), which is the main data used in this thesis. It is a machine readable corpus of English news reportage with discourse annotation. The profile and the illustration of the coding scheme of the AT, as well as some theoretical topics related to the coding scheme will be discussed.

Chapter 4 includes the observation of cataphora in the AT and other corpora with particular attention to common structural patterns and the possibility of anaphoric reading.

In Chapter 5, I will discuss the generative syntactic approaches to cataphora. 5-1 focuses on Reinhart's C-command coreferential constraints (C-command model), which is one of the most influential syntactic domain models. Its influence can be found not only in the generative framework but also in some cognitive approaches such as Ariel (1990) (discussed in 6-1).

5-2 deals with backwards pronominalization rules originally formalised by Langacker (1969) (often called 'Precede and Command' or 'Precede/Command'), and re-formulated by Carden (1986). After illustrating the models, I will discuss empirical problems of backwards pronominalization rules, and examine Carden(1982:379)'s claim that cataphora is a sentential phenomenon and "obeys a structural condition". In 5-3, I will discuss Binding theory (Chomsky, 1981, 1985, 1986) and cataphora. Although cataphora was one of the central issues in the generative approach in 1960's and 1970's, it has been almost neglected in Binding theory. I discuss the theoretical reasons for the neglect of cataphora along with the problematic examples to Binding theory.

Chapter 6 and 7 focus on how cataphora data can be accounted for in cognitive discourse theories (Mental representation theories).
In Chapter 6, I will first explore the hierarchical approaches to account for the relationship between the cognitive status of a discourse entity and various referring expressions, namely the 'Accessibility Scale' (Ariel, 1990) and the 'Givenness Hierarchy' (Gundel et al. 1993). I will examine Ariel's empirical predictions on cataphora against the AT data in 6-1, and I will compare the claim made by 'Givenness Hierarchy' (Gundel et al, 1993) with cataphora data in 6-2. I will then discuss Cornish (1996)'s Account on cataphora, which is a sceptical view on cataphora, in 6-3.
In 6-4, I deal with the theoretical implications of cataphora for the current mental representation theories. In 6-5, I will discuss a problem of current mental representation theories, namely a practice of equating 'referent' with 'mental

representation'.

In chapter 7, I will attempt to investigate how the notion of reference-direction (anaphora and cataphora) can be accounted for in on-line (sequential linear) text analysis. This attempt is made in order to examine the argument (made by Wales (1996) and Stockwell(1995)) that the notion of anaphora and cataphora can be dispensed with "virtually altogether" from the point of view of on-line text processing.

After explaining the mental representation (MR) theorists' view on the notion of reference-direction (anaphora/cataphora) in 7-1, I will take a preliminary look at the borderline cases of anaphora and cataphora observed in the Anaphoric Treebank (AT) in 7-2.

I will then discuss the following factors relevant to the reference-direction judgement in each section:

7-3. The initiation point and completion point of interpretation, and the gap between them.

7-4. Suspension of interpretation caused by structural dependency.

7-5. Immediate initiation and rapid completion of interpretation of pronouns

In 7-6, I will summarise the overall picture of reference-direction judgement observed in the AT, and discuss whether the notion of anaphora and cataphora can be dispensed with "virtually altogether" from the point of view of on-line processing, the argument made by Wales (1996) and Stockwell(1995).

Chapter 8 is the summary of discussions and conclusion.

Chapter 2: An Overview of Personal Pronoun Reference in New Reportage.

In this chapter, as a foundation for the discussion in the following chapters, I will overview the interpretation of a personal pronoun reference in particular focusing on human personal pronoun references in news reportage.

2-1. Some Features of News Reportage Texts Relevant to Personal Pronoun Reference.

A news reportage text usually employs an 'impersonal' style of narration, **third person narration,** "where reference by the narrator to himself is avoided" (Leech and Short, 1981: 266). Hence whereas there are many instances of first-person pronouns (*I, my, me myself*) in direct quotes, the narrative of news reportage text rarely includes first-person pronouns (1PPs), which refer to the writer him/herself.

Also as Leech and Short note, "the absence of an 'I' invites the reader to assume that there is no explicit 'you'" (ibid.). This tendency is supported by the Anaphoric Treebank (AT). The narrative in the Anaphoric Treebank (AT) does not have any instances of direct reference to the writer. There is only one instance of the direct reference to the readers:

[A003 1] Motorists: concentrate on the car, but keep **your** eye on the road.

In a transcription of a TV or radio news programme, on the other hand, the reporter's narration sometimes includes first-person pronouns (1PPs) referring to the reporter him/herself, as well as second-person pronouns (2PPs) referring to the addressee (interviewees), as in the cases of 1PP *I* and 2PP *you* (typed in bold) below:

> Mr. Fowler would only say "I have no comment to make." Well, **I** have the minister on the phone now:
> Mr. Fowler, do **you** still refuse to answer these criticisms?
> (change of speaker)
> What I said this morning, and I shall repeat it again now to you, was …
> (taken from Fligelstone, 1989)

The receivers (addressees) of a news report are usually infinitely many,

anonymous readers. As Leech and Short note:

> For all published texts, … there is usually one addresser but a large number of
> addressees, the vast majority of whom the writer has never met.
> (Leech and Short, 1981:258)

Since the vast majority of readers are unknown to the writer of an article, the

immediate situation shared between a writer and readers is very limited, though a

writer can still assume the background knowledge shared with the 'average' readers.

2-2. Deictic Dependency of Interpretation of Personal Pronoun Reference

A news reportage text often includes a 'direct-quote construction' (Givón,

1990:532), which consists of direct speech (DS) and a reporting clause attached to the

piece of DS. When we interpret a direct-quote construction, we need to take account

of the obvious fact that the addresser (speaker) of the direct speech (DS) is usually

different from the addresser (writer) of the news article who inserts the piece of DS in

the narrative.

For instance:

[A003 85] *Novelist Wright Morris* says *he* always is working on a book "because this
    is the way *I* breathe."

The discourse entity 'Wright Morris' is mentioned with the third person pronoun

(3PP) *he* by the addresser (writer) of the article, since for the writer (first person) and

the readers (second person) of the article, 'Wright Morris' is a third person entity. In the direct speech (DS), on the other hand, 'Wright Morris' is no longer mentioned with a third person pronoun (3PP) but with the first person pronoun (1PP) *I*, since the addresser of the piece of DS is not the writer of the article but 'Wright Morris' himself. When we resolve the personal pronouns in a direct speech (DS), first of all we need to take account of the point of view (POV) of the speaker of the piece of DS, as Short notes:

> In reported clauses using DS ... , all the linguistic features used must be related to **the speaker's viewpoint**. These could involve any of the viewpoint markers ...
> (Short, 1996:300)

The speaker's viewpoint is often termed **deictic center** (Segal, 1995:15, Brown and Yule 1983:52). The notion of **deictic center** or **'origo'** (in Bühler (1982)'s term) is concisely illustrated by Segal:

> At the center, or origo, are ,*here*, *now* and *I*. Something that is *here* may be *this*. *That* is not *here* but *there*. I address *you*, and refer to *him* or *her*. Both temporal adverbs (e.g. *then*, *soon*) and tense (present, past, future) mark the temporal relation to the origo.
> (Segal, 1995:15)

It can be said that a direct quote construction includes two distinct speakers' viewpoints (deictic centers), as Givón notes:

> Further, the points-of-view are now de-coupled: The speaker retains deictic responsibility for the main clause, while ceding responsibility for the complement clause to the subject of 'say'.
>
> ...
>
> The opening of a direct quotation is akin to initiating a new universe of discourse. The two events, coded in the main and complement clause, thus exist, in some sense, in two separate universes of discourse; although the universe of complement event is embedded within that of the main-clause event.
>
> (Givón, 1990: 531-532)

Hence when a reader proceeds from the narrative to a piece of direct speech (DS), s/he needs to take account of the shift of deictic center (often called a **deictic shift**), and interpret all the linguistic features related to deictic center accordingly. The linguistic features related to deictic shift include: tense, pronouns, other deictic markers besides tense and pronouns, speech act indicators, indications of voice quality and other phonetic aspects, and colloquial lexis (Short, 1996:300).

Obviously, the writer of the article is not responsible for coding the expressions of DS. Yet a writer can decide:

(i) the portion (or whole) of the original speech to be inserted.

(ii) the point in the narrative at which the extracted DS is inserted.

In other words, a writer has a kind of cut-and-paste control over DS.

A direct quote construction is the result of the writer's act of inserting a piece of DS into the narrative. Although the piece of DS is not coded by the writer, as a result of the insertion, a reader needs to resolve the pronouns in the piece of DS and the narrative consistently, taking account of the de-coupled points of view and deictic conditions.

Hence in analysing a text, it is convenient to make a distinction between **cross(inter)-deictic coreference**, i.e. a coreference between an NP in the narrative and an NP in a piece of DS, and **intra-deictic coreference**, i.e. a coreference between NPs within the narrative or within a piece of DS.

2-3. Situational Reference in Speech and Writing

It is often said that the referent of a personal pronoun is usually recoverable (accessible) either from the co-text or the immediate situation of the discourse. CGEL notes:

> Recoverability is crucial to the use of pro-forms. Pro-forms such as personal pronouns have very unspecific meanings ... Therefore it is necessary, for the interpretation of these words, to have information from which we can uniquely predict their intended referents. This information is usually found in a preceding or following part of the text, but it may also be found in the situation.
> (CGEL:862)

This section focuses on situational reference, where a pronoun is used to refer to an entity in the immediate situation, in speech and writing.

2-3-1. Situational Reference in Speech

In speech, even if the referent is not verbally mentioned before directly or indirectly, it can be referred to with a 3PP if the referent is recoverable (accessible) in the immediate situation shared between the speaker and the receiver, as in the following example cited from Brown and Yule:

(A large dog approaches A and B, A says to B:)

"I hope it's friendly."

(Brown and Yule, 1983:215)

In this example, the referent (the large dog) can be considered to be already salient in

the receiver(B)'s memory when the 3PP is used. In this respect the referent of the 3PP is not new information but given information in the discourse.

Yet in speech, even if the referent is new information in the discourse, i.e. it is not verbally mentioned before nor is salient in the receiver's memory, the referent can be referred to with a 3PP when it is recoverable in the shared immediate situation. For instance,

(The speaker is pointing to a man in the street and says:)

"Look at him !"

In this case, the referent can be never known to the receiver before (the referent can be 'brand new' in Prince (1981)'s term).

2-3-2. Situational Reference in Writing

In writing, on the other hand, the addresser and the receiver are usually separated in space and time. The shared immediate situation for the situational use of a 3PP is highly limited compared with that in speech.

Although very limited, however, there is an immediate situation shared between the writer and readers independently of any medium of writing, whether the text is written on paper or a computer screen. That is what might be called a 'cursor point', namely the point on a text at which the reader's attention is engaged. Whatever the form of writing, it can be assumed that the writer and readers always share the

cursor point as their immediate situation. The cursor point can be further extended to the currently engaged page/screen, or currently engaged document (either paper form or machine readable form), though somewhat less independently of written medium.

It is this shared immediate situation (cursor point, currently engaged page/screen, document) that the interpretation of the deictic expressions called 'discourse-deixis' (Levinson, 1983) is based upon.

For example, the interpretation of deictic expressions such as 'the diagram **below**', 'the figure **on the right**', or 'in the **next** page' necessarily assume the cursor point or currently engaged page. Theses deictic expressions can be interpreted only with reference to the shared immediate situation of writing[4].

If there is a visual source in the currently engaged page or screen, a 3PP reference is possible without having a verbal antecedent, as in the following example typically found in children's books or foreign learner's language text books:

(Assuming that there is only one male figure in the picture)

"Now, look at the picture, what is <u>he</u> doing ?"

---

[4] In the Anaphoric Treebank (AT), discourse deixis is treated as METATEXTUAL REFERENCE. (see Fligelstone, 1991)

In any case, because of the highly limited shared immediate situation, a 3PP reference is usually made with a coreferential or co-specifying full NP in writing. In other words, a 3PP tends to have a textual antecedent in writing[5], as Emmott notes:

> In a film, focusing could be achieved by pointing, but in non-visual medium of written narrative a verbal antecedent is normally required.
> (Emmott, 1997:217)

---

[5] The referent of a 3PP can be recoverable if it is already indirectly mentioned (indirect anaphora), such as:

> I went to a concert last night. **They** played Beethoven's ninth.
> (cited from Chafe, 1972)

According to Sidner, this kind of indirect anaphoric use of pronoun is typically found in speech rather than in writing:

> Such examples, as far as I can tell, do not occur naturally in written samples.
> (Sidner, 1983b:317)

2-4. Interpretation of Personal Pronouns in Direct Speech

2-4-1. Difficulties involved in Interpretation of Personal Pronouns in Direct Speech

A pronoun in a piece of direct speech (DS) is originally encoded by the speaker of the piece of DS taking account of the receiver's condition of memory and the shared immediate situation between the speaker and the receiver. The immediate situation, however, is distinct from the immediate situation shared between the writer and readers of the article. A reader of the article does not usually share the immediate situation with the speaker of the piece of DS. This gives difficulty in the interpretation of a pronoun in a piece of DS when the reference is situational.

Consider:

[A057 37] The primitive stone table and chairs where the treaty was signed in 1844 still sit at the same spot in the Chinese temple with the resonant name of Kun Lam Tong or goddess of mercy.
[A057 38] "What do you know," said a surprised elderly American tourist when the guide related the historic event.
[A057 39] "Not the easiest place to sign a treaty on, is **it**?" he said, ...

The DS in [A057 39], *"Not the easiest place to sign a treaty on, is it?"*, is addressed by the 'surprised elderly American tourist' to the listeners who shared the immediate situation with the American tourist. The listeners were assumed to be able to access the referent of the 3PP *it* in the immediate situation shared with the 'American tourist', and they were assumed to be able to resolve the 3PP *it*. Yet readers of this article do not share the original immediate situation of the piece of DS, and it is difficult for the readers to resolve the 3PP *it* in the piece of DS. The writer, therefore, needed to supply the extra information for a reader to resolve the 3PP:

[A057 39] "Not the easiest place to sign a treaty on, is *it*?" he said, running his hand over *the rough surface of the table*.

Writers of news report often supply extra information for readers to resolve a pronoun in a piece of DS in the following ways:

(i) Inserting extra information in the piece of DS using brackets:

[A059 20] "**They (Boston College)** executed well and got a lot of movement and we were just standing around " in the first half.

[A045 14] **It (the press)** did create some turnovers but the key thing was it spread out the defense," Westhead said.

(ii) Inserting extra information in the reporting clause using "said of" or "referring to":

[A006 32] "We're working hard to increase *it* to at least $840," he **said of** *the base salary*.

[A041 26] *He* has so much intestinal fortitude," said Malmquist, **referring to** *the top U.S. cross-country skier*.

[A048 22] *He*'s a hell of an athlete, " Knight **said of** *the Big Ten's No.2 scorer*.

The notion of first person, second person, and third person themselves are mutually exclusive, though plural pronouns in first and second person often involve inclusive references. The mutually exclusiveness of 'person' relations are usually realised in singular human personal pronouns. These inclusive/exclusive relations are clearly shown in the following table presented in Quirk et al, 1985 (CGEL).

s: the originator(s) of the message (speaker/writer)

h: the addressee(s) of the message (hearer/reader)

o: any other referent(s) excluded from the definition of s and h.

	s	**h**	**o**	*PERSON*	*EXAMPLES WITH REFLEXIVE PRONOUNS*
S	+	-	-	1st	I gave myself up.
i	-	+	-	2nd	You should be ashamed of yourself, Richard!
n	-	-	+	3rd	Mary has hurt herself.
g					John has hurt himself.
					The spider has hurt itself.
p	+	-	-	1ST	We, the undersigned, pledge ourselves to ...
l	+	+	-	1ST	We complemented ourselves too soon, John. ['inclusive we']
u	+	-	+	1ST	The children and I can look after ourselves. ['exclusive we']
r	+	+	+	1ST	You, Ann, and I are working ourselves to death.
a	-	+	-	2nd	You ought to be ashamed of yourselves, children!
l	-	+	+	2nd	You and John will have to cook yourselves.
	-	-	+	3rd	They helped themselves to coffee and cakes.

(taken from CGEL:430)

These mutually inclusive/exclusive referential relations of 'person' provide certain formal clues for resolving pronouns in direct speech, and for analysing cross-deictic coreferences.

First Person Pronouns (1PPs) in direct speech (DS)

A first person singular pronoun (1PSP) in a piece of DS usually has a positive coreferential relation with the NP referring to the addresser (writer) of the narrative (mostly realised as the subject of the reporting verb if it is not passive), whereas it has negative coreferential relations with the NPs referring to any other persons in the narrative. For example:

[A051 80] "**I** saw that it was a traumatic experience for *her* to see that stone," said **Mrs. Vrhovac**, who was with *Mrs. Kovatch* when *she* first saw the inscription.

The 1PP *I* in the piece of DS has a positive coreferential relation with the subject of the reporting verb *Mrs. Vrhovac*.

Second Person Pronouns (2PPs) in direct speech (DS)

A second person singular pronoun (2PSP) in a piece of DS has a positive coreferential relation with the NP referring to the receiver in the narrative, whereas it has negative coreferential relation with the NPs referring to any other persons in the narrative.

[A014 32] A flight attendant awaiting the same flight recognized him and told *the gate agent*, " Do *you* know we have one of the most famous college football coaches in America flying with us?"

The 2PP *you* in [A014 32] has a **positive** coreferential link with the **object** of the

reporting clause *the gate agent*, while it has a **negative** coreferential link with the subject of the reporting clause *A flight attendant*, which refers to the agent of the reporting clause.

A 2PP in a piece of DS can be coreferential with the subject of the reporting clause when the reporting clause is passive:

[A010 115] "Did *you* ever take a bribe?" *he* was asked.

Third Person Pronouns (3PPs) in direct speech (DS)

A third person human singular pronoun (3PHSP) in a piece of DS usually has negative coreferential relations with the NPs referring to the speaker or the receiver of the piece of DS in the narrative. A 3PHSP in a piece of DS does not have a positive coreferential relation with NPs in the narrative.

In [A051 80] below, the 3PP *her* in the piece of DS has a negative coreferential relation with the subject of the reporting clause *Mrs. Vrhovac*. There is no formal indicator of the positive coreferential relation of the 3PP, though the 3PP is coreferential with *Mrs. Kovatch*.

[A051 80] "I saw that it was a traumatic experience for **her** to see that stone," said
         Mrs. Vrhovac, who was with **Mrs. Kovatch** when **she** first saw the
         inscription.

Whether a 3PHSP in a piece of DS has a coreferential NP in the narrative depends on each case. The lack of positive coreferential relation in the cross-deictic

3PP coreference causes difficulty for pronoun resolution.

Within a piece of DS (or intra-deictically), some pronouns in a piece of DS can be resolved based on the ordinary syntactic coreferential preferences or syntactic parallelism:

A Syntactic coreferential preference

[A002 16] "The Klan isn't new, but we also have to recognize not all members of **the Klan** show **themselves** like they have here recently," said David Hill, a leader in the American Indian Movement.

Syntactic parallelism:

[A005 61] "**Ben** whetted my appetite.
[A005 62] **He** showed me how to be more patient," said the oft-frustrated Burns, who
...

2-5. Interpretation of Personal Pronouns in Narrative

In a piece of news reportage, the narrative rarely has a direct reference to the writer (1PP) and the readers (2PP) of the article. Hence only third person pronouns (3PPs) are relevant here.

Within a narrative (or intra-deictically), some pronoun references can be resolved based on the ordinary syntactic coreferential preferences or syntactic parallelism, in the same way as pronoun references in direct speech (DS):

A syntactic coreferential preference:

[A017 91] It said **Cone** should disqualify **himself** from the case ...

Syntactic parallelism:

[A002 20] **One speaker, Naomi Mendus**, said the rally was scheduled "to call for the outlawing for the Ku Klux Klan."
[A002 21] **She** said **she** feels there is ...

It is widely recognised among discourse theorists that one of the main functions of a 3PP is to signal to maintain the highlight (focus) placed on the previously focused discourse entity (highlight (focus)-maintenance). The remainder of this section deals with the topics related to triggering and maintaining 3PP highlight (focus).

2-6. Signalling highlight (focus) maintenance: a Contrastive function of Third person pronoun (3PP) in relation to full NP

In contrast with personal pronouns, proper names and indefinite or definite lexical NPs are often called full NPs. The notion of 'full NP' is useful in particular when we talk about the contrastive function that a 3PP has in relation to a proper name or a lexical NP. One such contrastive function (signalling highlight (focus) trigger and maintenance) is discussed in this section.

In writing, when a discourse entity (other than the addresser and the receiver) is introduced into the discourse for the first time, it is usually referred to by a proper name or a lexical NP. For instance:

[A014 30] ... *Ara Parseghian*, who coached for 11 years (1964-74) at a small Catholic school in South Bend, Ind., before *he* found *his* niche as a television color commentator, checked in last month for a flight at Chicago's O'Hare Airport, where *he* probably has spent a good part of *his* life.

The person 'Ara Parseghian' is highlighted by a proper name *Ara Parseghian* at the beginning of the sentence. Subsequent references to 'Ara Parseghian' made by 3PPs signal to a reader to maintain the highlight on the currently focused entity 'Ara Parseghian'.

In [A017 64] below, the use of a lexical NP *A London diamond dealer* highlights its referent, and the highlight is maintained by using the following 3PPs.

[A017 64] *A London diamond dealer* who said *he* was on *his* way to set up a deal with ...

The point when a person is first introduced in discourse with a full NP can be compared with the moment when a person on stage is highlighted for the first time. The subsequent reference to the person with 3PPs can be compared with the maintenance of the highlight upon the person.

This 'stage' metaphor, initially introduced by Chafe (1970, 1974), is used by cognitive discourse theorists for illustrating a reader's focus of attention on a particular discourse entity.

Emmott says:

> ... priming is like characters standing on a stage in front of us, whereas focusing is like someone drawing our attention towards one of these characters.
> (Emmott, 1997:217)

The phenomena related to 'focus of attention' has been approached from different theoretical points of view, and the terminologies and concepts to describe the phenomenon are diverse among the theories, though often overlapping each other. In this thesis, following Chafe and Emmott, I will use the straightforward 'stage' metaphor to describe the contrastive function of a 3PP in relation to a full NP, and use the term 'highlight' or 'focus' in the sense of a reader's focus of attention on a particular discourse entity 'on stage'.

We can assume a **male-highlight** and a **female-highlight** corresponding to the male 3PPs (*he, him, his, himself*) and the female 3PPs (*she, her* (objective), *her* (possessive), *herself*) respectively. Since it corresponds to singular forms, a male-highlight can be focused on only one person at a time, and the same is true for a

female-highlight. A male-highlight and a female-highlight can be maintained together so long as gender ambiguity is not involved. For instance:

[A003 108] **Dr. Steve Bramwell** of the University of Washington Sports Medicine Clinic said <u>Lloyd</u> had acute gastroenteritis, fever and severe dehydration. [A003 109] **He** said <u>she</u> should not play.

The male-highlight on 'Steve Bramwell' and the female-highlight on 'Lloyd' can be maintained without interfering with each other in [A003 109].

2-7. Re-focusing / Re-highlighting out-of-focused entity

The crucial point in highlight maintenance is that once a highlight is shifted from one person to another, the person who becomes out of focus usually cannot be re-highlighted (placed in focus) again by using a 3PP. Consider:

[A010 93] **Stevens** said reporters should not infer from **his** comments either that he did or did not know of other senators involved.

[A010 94] **He** said **he** had "asked for that information" but did not say whom **he** had asked.

[A010 95] <u>Democratic Leader Robert Byrd of West Virginia</u>, asked if <u>he</u> knew of any other senators under investigation, said <u>he</u> did not.

[A010 96] **Stevens** said **he** did not believe Republicans should use the investigation, which has so far focused mainly on Democrats, as a campaign issue.

The person 'Stevens' is highlighted with the main subject *Stevens* in [A010 93]. The highlight on 'Stevens' is maintained with 3PPs in [A010 93] and [A010 94]. The highlight, then, is shifted form 'Stevens' to 'Robert Byrd' by the use of the full NP *Democratic Leader Robert Byrd of West Virginia* (underlined) in [A010 95]. The highlight focused on 'Robert Byrd' is maintained by the 3PPs (underlined) in [A010 95]. Since the highlight is shifted to another person, the out-of-focus 'Stevens' cannot be put in focus again by using a 3PP. So if the full NP *Stevens* in [A010 96] is replaced with a 3PP *He* (underlined) as shown below ([A010 96']), the 3PPs *He* and *he* in [A010 96'] are likely to signal the maintenance of the previously highlighted 'Robert Byrd' rather than to refer to 'Stevens', which would result in incoherent discourse:

[A010 93] **Stevens** said reporters should not infer from **his** comments either that he did or did not know of other senators involved.

[A010 94] **He** said **he** had "asked for that information" but did not say whom **he** had asked.

[A010 95] *Democratic Leader Robert Byrd of West Virginia*, asked if *he* knew of any other senators under investigation, said *he* did not.

[A010 96'] <u>He</u> said he did not believe Republicans should use the investigation, which has so far focused mainly on Democrats, as a campaign issue.

In order to re-highlight the out-of-focus 'Stevens', a full NP form (in this case *Stevens*) is needed.

A similar observation is made by Brown and Yule (1983). They make a distinction between a 'current (given) entity', which is equivalent to a previously highlighted entity, and a 'displaced (given) entity', which is equivalent to an out-of-focus entity. Based on their diagram drawing data[6], they make the following observation:

> In this data, pronouns never occur as expressions identifying displaced entities, only as expressions identifying current entities.(p.174)
> ... displaces entities are regularly referred to by full lexical definite NPs.
> (Brown and Yule, 1983: 176)

In other words, Brown and Yule state that pronouns are used to maintain the highlight on previously highlighted entities; they are not used to highlight out-of-focus entities.

The point that a 3PP usually cannot highlight a discourse entity but maintains a previously highlighted entity is mentioned by several cognitive discourse theorists, such as Sanford and Garrod:

---

[6] See Brown and Yule, 1983: pp.172-174.

"Studies of production in naturalistic settings suggest that noun phrases are used to **reintroduce individuals** which have **slipped from focus**, whereas pronouns are used **to maintain reference to focused individuals** (e.g. Marslen-Wilson, Levy, & Tyler, 1982)."
(Sanford and Garrod, 1989:245 -- my emphasis)

Mckoon, Gerrig, and Green (1995) state that a pronoun "confirms accessibility that already exists":

A pronoun does not create accessibility for itself--it confirms accessibility that already exists. (Mckoon et al, 1996)

Their claim, "a pronoun does not create accessibility for itself", may be equivalent to stating that a 3PP usually cannot highlight a discourse entity for itself.

Green et al (1994) treat a pronoun as "a recall cue":

Successful resolution of a pronoun requires that some entity be sufficiently accessible in the comprehender's discourse representation to provide a unique match to the pronoun as **a recall cue**;
(Greene et al, 1994:513)

Cornish defines anaphora as "a signal to continue the existing attention focus already established" (Cornish, 1996:22), and states that 3PP signals the addressee to **maintain** the already focused entity:

The pronoun *he* ... presupposes that its referent is already salient, that is, 'in focus' within this stereotypical situation ... and prompts the addressee to maintain that high level of saliency
(Cornish, 1996:23)

Because of this functional aspect of 3PPs, i.e. signalling the maintenance of current

highlight, 3PPs are usually considered to be anaphoric by default.

A highlight maintenance signal is not supposed to precede a highlight trigger signal. If a pronoun is primarily a highlight maintenance signal, it has to be preceded by a highlight trigger signal, namely a full NP. If this is the norm, it is impossible to talk about cataphora, in which a pronoun (a highlight maintenance signal) precedes its antecedent full NP (a highlight trigger signal). The widespread scepticism towards cataphora (discussed in chapter 6) may be rooted in the fact that a 3PP cannot re-highlight/re-activate an out-of-focus discourse entity.

2-8. Grammatical Factors trigger highlight: A claim made by Centering theory.

To identify the factors which trigger highlight (focus) has been one of the central issues in discourse studies, and the details of the researches done on this issue cannot be covered in this thesis. In this section, I will briefly look at the claim made by **Centering theory** (Walker et al, 1998), which is one of the major theories dealing with this issue.

The discourse entities evoked by an utterance in a discourse segment are defined as a set of 'forward-looking centers (Cfs)' in Centering theory. They are ranked according to discourse salience (Cf ranking). According to Centering theory, the most salient entity of the utterance (termed as a 'preferred center' (Cp)) is likely to be realised as a pronoun or an elliptical form in the next utterance. In Centering theory's terms, the 'preferred center(Cp)' of an utterance is likely to be the 'backward-looking center (Cb)' of the next utterance (Walker et al, 1998:3). In Centering theory, therefore, the saliency ranking of the evoked entities (Cf ranking) of an utterance determines, to a large extent, what is likely to be pronominalised in the next utterance. Walker et al state that there are a number of factors which influence the saliency ranking (Cf ranking):

> The Cf ranking is determined by a number of factors, such as the grammatical role in which the entity is realised, surface order of realisation, and information status
> (Walker et al, 1998:4).

With regard to the grammatical role, Centering theory proposes the following order of

Cf Ranking by Grammatical Function:

Subject > Object > Other

(Brennan et al 1987, Walker et al, 1998)

This order claims that the discourse entity realised in the subject position is assumed to be more salient than the entity realised in the object position, and both entities (realised in subject and object position) are assumed to be more salient than the entities realised in any other grammatical functions. Hence according to Centering theory, as far as the grammatical role is concerned, the subject NP of a sentence is the most likely highlight(focus)-trigger.[7]

This claim is supported by Mitkov's empirical observation made on a computer science texts corpus[8]:

> ... we found that the subject is the primary candidate for center (in about 73% of the cases). The second most likely center would be the object (25 %) and the third most likely one the verb phrase as a whole (2%).
> (Mitkov, 1994:151)

---

[7] Because of the limitation of time, I could not test the Centering theory's claim against the Anaphoric Treebank (AT).

[8] Yet Mitkov also reports that "there are many additional interrelated factors" (such as verb semantics, repeated concepts, headlines) "which influence the location of center" and that "general preference for subject or object is not very accurate". (Mitkov, 1995)

2-9. Highlight maintenance at different levels

A significant feature of 3PP highlight (focus) maintenance in news reportage texts is that a highlight (focus) associated with a **reporting** part of a text can be maintained with a highlight associated with the **reported** part of the text simultaneously.

Consider:

[A001 121] **Ms. Cerino** said <u>the woman</u>'s acknowledgement of the serious nature of <u>her</u> offense will help significantly in the rehabilitation process.
[A001 122] **She** says the realization on the part of the offender that crime is wrong is often the key to rehabilitation.

Notice that between *Ms. Cerino* in [A001 121] and 3PP *She* in [A001 122] (both are typed in bold), there is a female 3PP (a highlight maintainer) *her*, whose antecedent (a highlight trigger) is *the woman* (both are underlined). This indicates that the female-highlight (focus) triggered by *Ms. Cerino* can be maintained without being affected by the local female-highlight (focus) which is triggered by *the woman* and maintained by the 3PP *her*.

It can be said that the cross(inter)-sentential female-highlight belongs to a different dimension from that of the local female-highlight (focus), and both female-highlights (foci) can be maintained together. Notice that the NPs associated with the cross(inter)-sentential female-highlight are the subject of the reporting verbs (*Ms. Cerino said; She says*).

The same point can be made for the following example:

[A047 43] "They made it seem like he wouldn't win if he didn't come home with five golds," says **Bob Brennan** of the local Chamber of Commerce.

[A047 44] But another factor, he notes, is that people here, unlike most of the rest of the country, are familiar with <u>Heiden</u>'s unprecedented dominance of world speedskating since <u>he</u> captured <u>his</u> first world championship three years ago.

[A047 45] *He* points out that many townfolk were angry at one pre-Olympic ABC report which indicated many Madisonians didn't even know who Heiden was, and the city council filed a formal protest with the network.

The cross(inter)-sentential male-highlight placed on 'Bob Brennan' (realised in bold-typed NPs) can be maintained without being affected by the local male-highlight (focus) placed on 'Heiden' (realised in underlined NPs). The NPs associated with the cross(inter)-sentential male-highlight are also the subject of the reporting verbs (*says Bob Brennan*; *He points out*).

The following example shows that the female-highlight placed on the entity in direct speech (realised in underlined NPs below) can be maintained independently from the female-highlight placed on the entity in the narrative (realised bold-typed NPs below):

[A015 47] **Ms. Hodge** said **she** became hysterical upon realizing that Robert had raced upstairs in search of <u>Mrs. Dumas</u> in <u>her</u> bedroom.

[A015 48] "He was looking for <u>her</u>, and he didn't want to leave without <u>her</u>," **Ms. Hodge** said.

[A015 49] "They went everywhere together, including shopping and to church every Sunday.

[A015 50] He really loved <u>her</u>.

[A015 51] Several weeks ago, he said, If anything happens to <u>grandma</u>, the Lord can take me, too."

[A015 52] **She** said the boy had trouble sleeping at night unless he knew Mrs. Dumas was near him.

Listing also has the same effect as direct speech (DS). Consider:

[A030 68] **Beaumont** drew up a list of proposed changes and gave a copy to all of the 73 members who attended this week's IOC session in Lake Placid, prior to the XIII Olympic Winter Games.

[A030 69] This is what **he** wants to do:

[A030 70] - Do away with all national flags and anthems.

[A030 71] Have the Olympic flag hoisted and the Olympic Hymn played at every victory ceremony.

[A030 72] - Do away with the opening parade on a country-by-country basis.

[A030 73] Instead, have the athletes march into the stadium by sports, nation mingling with nation.

[A030 74] - Have the national flags of the competing countries grouped in the center of the field, away from the athletes.

[A030 75] - Have the President of the IOC declare the Games open, not a head of state or his deputy.

[A030 76] - Take all team sports - soccer, field hockey, basketball, handball - out of the Games, thus reducing what **he** says is an overcrowded program.

Despite seven intervening sentences, the writer seems to assume that the male-highlight placed on 'Beaumont' in [A030 68] still can be maintained in [A030 76], as the use of 3PP *he* in [A030 76] indicates.

Notice that in the intervening sentences, there is a local male-highlight in [A030 75] (triggered by the NP *a head of state* and maintained by the 3PP *his*).

These examples indicate that in news reportage, we need to assume a different level of 3PP highlight (focus) maintenance associated with **reporting** and **reported** part of the text.

2-10. Conceptual issues related to (re-) focusing (highlighting)

2-10-1. Proper Name in text and name (address) to be remembered

In on-line text analysis[9], it is important to distinguish between (i) a particular full NP in a text and (ii) a full NP form associated with the discourse entity in a reader's longer-term memory.

In order to re-highlight (re-activate) an 'out-of-focus' entity, a reader needs to identify the entity among the out-of-focus competing referent candidates. Also a reader needs to know that the referent is not a new entity but is already introduced in the discourse.

In practice, a writer may frequently, but need not necessarily, uses lexical repetition for subsequent non-pronominal references. A writer tends to use full NP forms which is usually less specific than an initial mention full NP. For instance:

[A010 88] **Acting Senate Republican Leader Ted Stevens** declined today to say whether more than one senator is under investigation in a Capitol Hill payoff scandal.
...
[A010 94] **He** said **he** had "asked for that information" but did not say whom **he** had asked.
[A010 95] Democratic Leader Robert Byrd of West Virginia, asked if he knew of any other senators under investigation, said he did not.
[A010 96] **Stevens** said **he** did not believe Republicans
[A010 97] "**I** am not going to make it into a political thing," **he** said.
...
[A010 101] Byrd said he did not have enough information to comment on the investigatory techniques.

---

[9] On-line text analysis concerns how a text is interpreted by human processor as the text unfolds (on-line). See 3-6 for further discussion of on-line and off-line analyses.

'Ted Stevens' is initially mentioned with full name *Ted Stevens* in [A010 88], and then subsequently mentioned by less specific form *Stevens* when it is re-highlighted in [A010 96]. 'Robert Byrd' is initially mentioned with full name *Robert Byrd* in [A010 95] and subsequently mentioned by less specific form *Byrd* when it is re-highlighted in [A010 101].

Lexical repetition usually assumes readers' capability of remembering lexical forms. So far as repetition of a proper name is concerned, when a writer repeats a proper name to mention an out-of focus discourse entity, s/he usually assumes that readers somehow remember the proper name associated with the entity and can recall the entity. The remembered name, whether it is remembered clearly or vaguely, is carried forward and utilized when readers try to re-activate the entity when it is out-of-focus.

In on-line text processing, as Emmott notes, a particular full NP in a text "focuses the reader's attention on the appropriate entity representation and once it has done this it plays no further role" (Emmott, 1997:224). Yet a full NP form associated with the entity in reader's longer-term memory still has a role to play; it is used to focus the reader's attention on the out-of-focus entity again, in particular when there are competing highlight candidates. Hence in on-line text analysis, it is important to be aware of the distinction between (i) a particular proper name in a text and (ii) a proper name form associated with the discourse entity in a reader's longer-term memory.

2-10-2. Definition of antecedent and anaphor

It is important to be aware that the antecedent/anaphor relation can be defined at least in three distinct contexts. It can be defined: (i) in the (initial/subsequent) mention relations, (ii) in highlight (trigger/maintenance) signal relations, and (iii) in regular coreferential relations.

2-10-2-1. Defining antecedent-anaphor relation in terms of initial mention/ subsequent mention relation

Consider the following text fragment:

(NOTE: The first sentence [A010 88] appears at the beginning of a text. Sentences from [A010 89] to [A010 92] are omitted since they are not relevant to the discussion here.)

[A010 88] **Acting Senate Republican Leader Ted Stevens** declined today to say whether more than one senator is under investigation in a Capitol Hill payoff scandal.
...
[A010 93] **Stevens** said reporters should not infer from **his** comments either that he did or did not know of other senators involved.
[A010 94] **He** said **he** had "asked for that information" but did not say whom **he** had asked.
[A010 95] Democratic Leader Robert Byrd of West Virginia, asked if he knew of any other senators under investigation, said he did not.
[A010 96] **Stevens** said **he** did not believe Republicans should use the investigation, which has so far focused mainly on Democrats, as a campaign issue.

The discourse entity 'Ted Stevens' is initially mentioned by the full NP

"*Acting Senate Republican Leader Ted Stevens*" in [A010 88], and then subsequently

mentioned by the 3PPs and the coreferential full NPs in the following sentences. It

can be said that the initial-mention full NP (initiator) establishes an information

gathering point (a new node) for the discourse entity ('Ted Stevens'), to which all the subsequent inputs about the discourse entity are linked. The initiator establishes a new node and all the subsequent NPs (full NPs or 3PPs) point to the already established node. Wales notes:

> ... semantically once a text or discourse is initiated which has a consistent topic for reference, then both the NP and the 3PP, and any other equivalent referring expression, will be all 'given' information and equal in informational content ...
> (Wales, 1996:36)

Here, Wales defines new vs. given information in terms of initial vs. subsequent mention. If we define antecedent vs. anaphor in terms of initial mention (establishing a discourse representation) vs. subsequent mention (pointing to the already established discourse representation) relation, then only the initial mention NP can be the 'antecedent', and all the subsequent mention NPs (full NPs or 3PPs) are defined to be 'anaphors'.

2-10-2-2. Defining antecedent-anaphor in terms of highlight trigger/maintenance relation

In terms of highlight trigger vs. maintenance relations, it can be said that *Stevens* in [A010 93] below triggers highlight on the discourse entity 'Stevens' and the following 3PPs *He* and *his* in [A010 93] and [A010 94] maintain the highlight:

[A010 93] **Stevens** said reporters should not infer from **his** comments either that he did or did not know of other senators involved.
[A010 94] **He** said **he** had "asked for that information" but did not say whom **he** had asked.
[A010 95] Democratic Leader Robert Byrd of West Virginia, asked if he knew of any other senators under investigation, said he did not.
[A010 96] **Stevens** said he did not believe Republicans should use the investigation, which has so far focused mainly on Democrats, as a campaign issue.

51

Hence it can be said that the full NP *Stevens* is the 'antecedent' of the following 3PP anaphors in [A010 93] and [A010 94]. In the next sentence [A010 95], the highlight is shifted from 'Stevens' to 'Robert Byrd'. The highlight is then shifted back to 'Stevens' in [A010 96]. There is a discontinuity of the highlight on 'Stevens' between the first two sentences ([A010 93] and [A010 94]) and the last sentence [A010 96]. Hence it can be said that the full NP *Stevens* in [A010 96] serves to **re-highlight** the discourse entity 'Stevens' and the following 3PPs *He* in [A010 96] serve to maintain the highlight. So in terms of highlight signal relations, not only the initial mention full NP but also the subsequent 're-highlighting' full NP can be said to be the 'antecedent' of the following anaphors (3PPs). In highlight signal relations, an antecedent (full NP) is a highlight-trigger signal and an anaphor (3PP) is a highlight-maintenance signal.

2-10-2-3. Defining antecedent-anaphor in terms of regular coreferential relation

Consider:

[A027 95] **The 21-year-old from the Austrian Tyrol town of Flickenberg** won the
men's downhill at the XIII Winter Olympic Games Thursday.
[A027 96] **He** calls **himself** a fighter, and there were many battles before this final
victory.

In terms of highlight trigger/maintenance relation, the NP *The 21-year-old from the Austrian Tyrol town of Flickenberg* in [A027 95] is the antecedent, and both 3PPs *He* and *himself* in [A027 96] are anaphors.

Yet according to syntactic constraints, the 3PP reflexive *himself* in [A027 96] has to be coreferential with the clause subject 3PP *He*. Hence if we define antecedent vs.

anaphor in terms of regular coreferential relations, the 3PP *He* can be said to be the 'antecedent' of the 3PP *himself.*

Also consider:

[A027 115] **(183 Stock 183)** said <**REF=183 he** had absolutely no problems on the course.

[A027 116] "<**REF=183 I** was good at the start and in the technical part at the top of the course," <**REF=183 he** said.

In terms of the highlight signal relation, the full NP *Stock* in [A027 115] is the highlight-trigger, hence 'antecedent', and the 3PP *he* in [A027 116] is the highlight-maintenance signal, hence an anaphor. In terms of the regular coreferential relation, the 3PP *he* in [A027 116] can be said to be the 'antecedent' of 1PSP *I* in the initial DS, since the singular 1PP *I* in a direct speech (DS) is usually coreferential with the subject of the reporting clause.

Thus it is possible to talk about antecedent vs. anaphor relation, at least, in terms of (i) initial-mention/ subsequent-mention relation, (ii) highlight-trigger / highlight-maintenance relation, and (iii) regular coreferential relation.

In regular coreference relations, a 3PP can be an antecedent, as shown in the examples above. In highlight (focus) relation, on the other hand, a 3PP cannot be an antecedent, since it usually cannot highlight a discourse entity, as discussed before (-> 2-7). In initial-mention/ subsequent-mention relations, there are at least two cases in which a 3PP can be the antecedent (initial mention): one is the use of pronouns in a stylistic technique called *'in medias res'*, which positions readers "as already 'in-the-know', even though they are not really" (Short, 1996:267), causing readers "to feel

intimately involved with what is going on at the beginning of a story" (ibid.); the other is first-mention cataphora:

[A033 32] Amid demands for *his* resignation, **Democratic state Rep. Russell J. Reynolds** apologized Friday for using a racial slur in a response to a survey.

I will look at the first-mention cataphora observed in the AT in chapter 4.

2-11. Summary of chapter 2

I will summarise the main points of chapter 2:

1. In resolving the personal pronouns in a direct-quote construction, it is necessary to take account of the shift of the point of view (POV) of the addresser (**deictic shift**).

2. The mutually inclusive/exclusive referential relations of 'person' provide certain formal clues for resolving pronouns in direct speech.

3. A contrastive function of a third person pronouns (3PP) in relation to a full NP is to serve to maintain the highlight (focus) placed on the previously highlighted discourse entity. Because of this function, third person pronouns (3PPs) are usually considered to be anaphoric by default.

4. The widespread scepticism towards cataphora may be rooted in the fact that a 3PP cannot re-highlight/re-activate an out-of-focus discourse entity.

5. In news reports, we need to assume a different level of 3PP highlight (focus) maintenance, presumably according to speech events /speech levels.

6. In on-line text analysis, it is important to distinguish between (i) a particular full NP in a text and (ii) a full NP form associated with the discourse entity in a reader's longer-term memory.

7. It is possible to talk about antecedent vs. anaphor relation, at least, in terms of (i) initial-mention/ subsequent-mention relation, (ii) highlight-trigger / highlight-maintenance relation, and (iii) regular coreferential relation.

Chapter 3: Description and Illustration of the Anaphoric Treebank.

The main corpus used in this thesis is called **the Anaphoric Treebank (The AT)**. It is a machine readable corpus of written English with the annotation of the anaphoric relations in the texts. This chapter illustrates the profile and the coding scheme of the AT, and discuss some theoretical topics related to the AT coding scheme.

3-1 A Profile of the Anaphoric Treebank (AT)

The **Anaphoric Treebank** (AT) is a machine readable corpus of written English in which the cohesive relations in the texts are annotated by means of **a cohesion mark-up scheme** developed by **UCREL**[10] in Lancaster University collaborating with **IBM T.J.Watson Research Center, Yorktown Heights, New York**.

The Anaphoric Treebank was made on the basis of a collection of news reports originally published as Associated Press newswire, called the **AP corpus** (see Leech and Garside, 1991).

The original motivation for constructing this corpus was "to investigate the potential for developing a probabilistic anaphor resolver." (Fligelstone, 1990). As anaphor resolution is one of the central issues in NLP (natural language processing), it was felt to be useful to create a corpus in which the texts are marked up to show explicitly the anaphoric relationships found in the texts. In late 1989, an agreement was made between the UCREL and Yorktown Heights teams, with funding from IBM, to construct a corpus marked to show a variety of anaphoric or, more generally, cohesive relationships in texts.

---

[10] UCREL (Unit for Computer Research on the English Language) comprises members of the Departments of Computing and of Linguistics and Modern English Language in Lancaster University. Since 1980, one of the main research goals of UCREL has been the creation of annotated corpora.

3-2 An Illustration of the Annotation Process.

The basic principle of encoding anaphoric relations is that, in a typical anaphoric relation between a pronoun (anaphor) and a non-pronominal NP (antecedent), the antecedent NP is enclosed in brackets and is given an index number which is unique to its referent within the text, and the pronoun is preceded by a symbol indicating the referential link to that numbered NP.

Let us look at how an anaphoric relation in the following portion of a text is encoded:

> Anything <u>Kurt Thomas</u> does, <u>he</u> does to win. Finishing second, <u>he</u> says, is like finishing last.

As underlined in the example above, 'Kurt Thomas' is first mentioned by a full NP *Kurt Thomas*, and subsequently mentioned by third person pronouns (3PPs) *he*. In the annotation process, the antecedent NP *Kurt Thomas* is enclosed in brackets and given an index number (e.g. '108'), and the 3PPs are preceded by symbols ('<REF=') with the same index number ('108') indicating the referential link to the full NP *Kurt Thomas*. The result of the annotation is shown below (annotations are typed in bold):

> Anything **(108 Kurt Thomas 108)** does, **<REF=108** he does to win. Finishing second, **<REF=108** he says, is like finishing last.

Before the anaphoric relationships were analysed and encoded, each text already included the following annotations:

(i)   A reference code for each sentence.

(ii)   A part-of-speech indication for each word (a part-of-speech tag)

(iii) An indication of the main constituent structure for each sentence. (parsing labels)

In the example above, each sentence is given a unique reference code ([A009 90] and [A009 91]), as shown below:

[A009 90] Anything Kurt Thomas does, he does to win.
[A009 91] Finishing second, he says, is like finishing last.

Each word in a text is given an appropriate part-of-speech (POS) tag by the **CLAWS** POS tagging system[11]. Each POS tag is selected from a set of approximate 170 tags. Each word and its POS tag are separated by an underscore "_" (POS tags are typed in bold below):

[A009 90] Anything_PN1 Kurt_NP1 Thomas_NP1 does_**VDZ** ,_, he_**PPHS1**
          does_**VDZ** to_**TO** win_**VV0** ._.
[A009 91] Finishing_**VVG** second_**MD** ,_, he_**PPHS1** says_**VVZ** ,_, is_**VBZ** like_**II**
          finishing_**VVG** last_**MD** ._.

In addition to the POS tags, a sentence includes the result of syntactic parsing (an indication of the main constituent structure of the sentence) in the form of a labelled bracketing notation called '**skeleton parsing**'[12] (Parsing labels are typed in bold below):

---

[11] For the CLAWS POS tagging system, see Garside et al (1987), chapters 3 and 4.
[12] Essentially, this is a phrase structure parsing scheme. The skeleton parsing of the AP corpus was done manually with 17 non-terminal symbols. The AP corpus was skeleton parsed by the UCREL team over the period 1988 to 1989.
(see Leech and Garside, 1991; Garside et al, 1993).

[A009 90] [N Anything_PN1 [Fr [N Kurt_NP1 Thomas_NP1 N] [V does_VDZ V] Fr] N] ,_, [N he_PPHS1 N] [V does_VDZ [Ti to_TO win_VV0 Ti] V] ._.

[A009 91] [Tg Finishing_VVG second_MD Tg] ,_, [Si [N he_PPHS1 N] [V says_VVZ V] Si] ,_, [V is_VBZ [P like_II [Tg finishing_VVG last_MD Tg] P] V] ._.

The following example shows the result of encoding the sentence references, the part of speech tags, the skeleton parsing labels, and the anaphoric annotations:

[A009 90] [N Anything_PN1 [Fr (108 [N Kurt_NP1 Thomas_NP1 N] 108) [V does_VDZ V] Fr] N] ,_, <REF=108 [N he_PPHS1 N] [V does_VDZ [Ti to_TO win_VV0 Ti] V] ._.

[A009 91] [Tg Finishing_VVG second_MD Tg] ,_, [Si <REF=108 [N he_PPHS1 N] [V says_VVZ V] Si] ,_, [V is_VBZ [P like_II [Tg finishing_VVG last_MD Tg] P] V] ._.

The original AP corpus was divided into units of approximately 100 sentences[13], and the syntactic and anaphoric markings were carried out on each of these units, so that the anaphoric reference numbering began afresh with each unit.

---

[13] As mentioned in Garside et al (1997:73), occasionally a text sample does not include the beginning of the original text.

3-3 An Overview of the mark-up of the Anaphoric Treebank.

The initial coding scheme was created by Geoffrey Leech (Lancaster University) and Ezra Black (IBM). The coding scheme was then elaborated and tested by applying it to corpus texts by the UCREL team, whose feedback triggered further elaboration and testing for the scheme. This development cycle was iterated several times.

The AT coding scheme inherits some part of the framework from the cohesion mark-up scheme presented by Halliday and Hasan (1976). Yet the resulting coding scheme differs, to a large extent, from its predecessor.

There are several factors that influenced the resulting AT coding scheme.

Firstly, for any kind of corpus annotation, it is important that the annotation needs to be carried out systematically and consistently on the whole corpus, avoiding as much as possible conflicting encoding caused by the different views among the team analysts.

Compared with syntactic analyses (parsing), however, the discourse level analyses of anaphoric relations involves much more interpretative process, and the possibility of disagreement in interpretation among analysts is much greater than syntactic analyses, as Fligelstone notes:

> The nature of the task, with its heavy reliance on interpretation, suggests that it may prove impossible to achieve such a high degree of inter-analyst consistency as with the parsing scheme ...
> (Fligelstone, 1991)

Because of the nature of the task, a decision was made to avoid too detailed a level of analysis.

Secondly, since one of the motivations to construct this corpus was "to investigate the potential for developing a probabilistic anaphor resolver" (Fligelstone, 1992), the corpus needed to include a substantial volume of annotated texts for exploiting statistical data from it. This caused the requirement of the speed of annotation, which became another reason for avoiding a too-detailed level of analysis.

Thirdly, although the coding scheme is influenced by Halliday and Hasan (1976), the resulting corpus was hoped to be theoretically neutral as much as possible, so that the corpus could be used by a wide range of researchers who do not necessarily share theoretical positions with each other. This became a further reason for avoiding too detailed a level of analyses.

It is natural, on the other hand, that the corpus is hoped to include information which meets a wide range of users' needs and theoretical interests.

The resulting scheme, therefore, reflects the resolution of the 'tension' between the practical requirement to avoid too-detailed a level of analysis (caused by the inter-analyst consistency, the speed of marking-up, and the demand for theoretical neutrality) and the requirement to meet potential users' theoretical interests as much as possible.

In order to facilitate the task of annotating texts, a task-oriented editor **Xanadu** was created as an interface between the analysts and the texts to be annotated. The code for this was written by Roger Garside of Lancaster University Computing Department.(See Garside (1993) for further information about **Xanadu**.)

3-4 An Illustration of the AT Coding Scheme Relevant to Personal Pronoun References

3-4-1. Types of cohesive relationship covered in the Anaphoric Treebank coding scheme.

The mark-up scheme of the AT (The AT coding scheme) covers the following types of cohesive relationship:

(i)   Proform reference (coreferential)

(ii)   Non-pronominal NP co-reference

(iii)   Proform reference (substitutional)

(iv)   Indirect definite NP anaphora

(v)   Links expressible as 'inferable of-complementation' of one non-pronominal NP by another

(vi)   Textually recoverable ellipsis

(vii)   Non-pronominal predications (copular relationship)

(viii) Metatextual reference or what is called 'Discourse deixis' (Levinson, 1983)

In what follows, I will briefly overview those properties of the coding scheme relevant to personal pronoun reference, namely (i) and (ii) in the above list. Further details of the notations and the linguistic background about (i) and (ii) can be found in Fligelstone (1991, 1992). Also for demonstrative pronoun references and the other types of cohesive relationship (from (iii) to (viii)), see Fligelstone (1991, 1992).

3-4-2. Identifying Textual Elements in Cohesive Relations.

3-4-2-1. Pronouns

Since personal pronouns are usually one word long, they can be identified by the immediately preceding type-symbols '<REF='.

For a compound coreferential proform such as "each other", extra encoding (a colon plus the number of words) is given:

> [A003 63] ... there's another reason for cutting the antlers off (82 the domesticated elk 82) - to prevent <REF=82 them from goring <REF=82:2 each other during the mating season.

3-4-2-2. Non-pronominal elements

A non-pronominal element (not necessarily an NP but a clause or other type of constituent) in a cohesive relation is usually bracketed and indexed:

> [A001 69] (71 University of Oregon football players used athletic department telephones last fall for out-of-state calls to parents and girlfriends in apparent violation of rules on financial assistance 71), the Eugene Register-Guard said Saturday.
> [A001 70] Head Coach Rich Brooks acknowledged (71 the practice 71) was ...
> ...
> [A001 72] "... I'm sure <REF=71 it goes on at other schools, ..."

A non-pronominal element is usually marked as a complete constituent (i.e. a non-terminal node which can be defined in terms of the skeleton-parsing scheme), except the following cases:

(i) Non-restrictive postmodification of NPs is ignored:

[A004 57] **(57** Stromberg **57),** who was a member of Norway's national ski jumping team until **<REF=57** he suffered a broken back in a meet three years ago, finished fifth in qualifying for four spots on the 1980 Norwegian Olympic team.

In [A004 57] above, the non-restrictive postmodifier *"who was a member ... three years ago"* is not recognised as a part of the antecedent NP.

(ii) Where an NP contains **self-referring anaphora**, only a sensible 'core' is marked:

[A049 93] **(181** A man **181)** who turned in **<REF=181** his girlfriend as an illegal alien after a lovers' tiff faces kidnapping charges and his being held in jail in lieu of $50,000 bail.

The restrictive postmodifier *"who turned in <**REF=181** his girlfriend ... a lovers' tiff"* is not recognised as a part of the antecedent NP since it contains the **self-referring anaphora** 3PP *his*.

(iii) A nominal entity in a premodifying position, though not a complete constituent, is marked if it has coreferential links to other NPs:

[A001 46] Six persons were treated for smoke inhalation at **(60** Weld County General Hospital **60)** and released, a **(60** hospital **60)** spokesman said.
...
[A001 67] A **(60** hospital **60)** spokeswoman said (64 another female student 64) cut <REF=64 her...

(iv) Complementizers such as 'that' are disregarded in clausal antecedents.

3-4-3. Marking Direction of Reference between Antecedent Non-Pronominal NP and Pronouns.

In a pronoun reference, if the antecedent NP is conceived to precede the pronoun, the pronoun is identified as **anaphoric** and the direction marker '<' is encoded before the type-marker 'REF'. If the antecedent NP is conceived to follow the pronoun, the pronoun is identified as **cataphoric** and the direction marker '>' is encoded before 'REF'. In the following sentence occurring at the beginning of a text, the initial 3PP *he* is **cataphorically** ('>') linked with the NP *"California Gov. Edmund G. Brown Jr.",* with which the second 3PP *his* is **anaphorically** ('<') linked.

> [A011 57] Predicting >REF=102 he will fare "much better than anyone expects" in Sunday's Maine caucuses, **(102** California Gov. Edmund G. Brown Jr. **102)** Tuesday compared <REF=102 his Democratic presidential rivals to "two peas in a pod."

When the reference-direction of a pronoun reference is difficult to decide, an indeterminate marker '?<>' is encoded before 'REF':

> [A021 48] Coach Harry Neale was impressed with his old stomping grounds Tuesday night but had mixed reviews about **(51** his Vancouver Canucks **51)**.
> [A021 49] "That was one of the worst games ?<>REF=51 we've played since ?<>REF=51 we left Vancouver," Neale said after **(51** the Canucks **51)** earned a 5-5 tie with the Hartford Whalers in National Hockey League action.

The antecedent of the plural pronouns *we* in [A021 49] is judged to be indeterminate; it can be either the NP *"his Vancouver Canucks"* in the previous sentence (i.e. the 1PPs are anaphoric) or the following NP *"the Canucks"* (i.e. the 1PPs are cataphoric).

3-4-4. Indexing: Identification of the Referent (Discourse Entity) of a Textual Element.

3-4-4-1. When the referent is a single discourse entity

In the following example, both the NP *Danz* and the pronoun *he* have the same referent that is a single discourse entity, indexed '29':

[A003 16] (**29** Danz **29**) said <REF=**29** he was against ...

3-4-4-2. When the referent consists of multiple discourse entities

Pronouns

In [A017 16], the 3PP *them* has two referents, 'Brown' (indexed '1') and 'other players' (indexed '12'), hence both index numbers are encoded after the type-marker '<REF=':

[A017 16] (**1** Brown **1**) said Molinas merely took <REF=**1** him and (**12** other players **12**) to restaurants and bought **<REF=1,12** them food, ...

Non-pronominal NPs

In [A026 7], the NP *the two companies* has two referents, 'Chrysler Corp.' (indexed '9') and 'its Japanese affiliate' (indexed '10'), hence both index numbers are attached to the brackets:

[A026 7] (**9** Chrysler Corp. **9**) may help (**10** its Japanese affiliate **10**) bring its cars up to date, another sign of closer relations between (**9,10** the two companies **9,10**).

3-4-4-3. When the referent is either one or the other discourse entity.

Pronouns

The referent of 3PP *They* in [A050 97] may be either 'A total of 17 players' (indexed '123') or '15 early arrivals' (indexed 124) but it is difficult to decide which. Hence both index numbers separated by a slash ('/') are encoded after the type-marker (REF=):

[A050 97] **(123** A total of 17 players **123)** joined **(124** 15 early arrivals **124),** who had been in camp since last week.
[A050 98] **<REF=123/124** They included newcomers pitchers Tom Underwood and Rudy May and catcher Rick Cerone.

Non-pronominal NPs

In the following example, the referent of the NP *the entire region* is difficult to decide. Since it is judged uncertain whether the co-referent is 'Central America' (indexed '12') or 'Central America and the Caribbean' (indexed '8'), both index numbers separated by a slash ('/') are attached to the brackets:

[A036 4] Mexico's oil riches are being used to expand the nation's influence in **(8 (12** Central America **12)** and the Caribbean **8).**
...
[A036 8] His activist stance for Mexico has not conjured up images of the imperialist "colossus of the north," **(12** Central America **12)**'s traditional view of Mexico since the 1820s when it tried to absorb **(8/12** the entire region **8/12)** after winning independence from Spain.

3-4-4-4. When the referent is not quite certain.

In some cases there is uncertainty about a reference, but there is a judgement that a selected referent is the most likely one.

Pronouns

The referents of the 3PPs *they* and *their* in [A007 18] are felt to be probably 'Phil Esposito' (indexed '28') and 'Ron Greschner' (indexed '29') but this is not quite certain, hence an uncertain indicator '?' is marked before each index number:

[A007 17] Right wing (27 Anders Hedberg 27) withdrew because of shoulder problems and was replaced by center (28 Phil Esposito 28), while defenseman (29 Ron Greschner 29) took over for teammate (30 Barry Beck 30) (elbow).
[A007 18] Monday, <REF=?28,?29 they devoted much of <REF=?28,?29 their time to the tenure of Alan Eagleson, executive director ...

Non-pronominal NPs

The referent of the NP *The acting GOP leader* in [A010 100] is felt to be probably 'Stevens' (indexed '76') but this is not quite certain; hence an uncertain indicator '?' is marked before the index:

[A010 99] (76 Stevens 76) called Williams "a close personal friend" and said: "I still don't believe it and won't [[Its better to use the normal spelling here - perhaps you should add a footnote to explain this.]] believe it until a jury returns a verdict, and I deep down don't believe that will happen."
[A010 100] (?76 The acting GOP leader ?76) said questions about the ...

3-4-5. Other Markings Relevant to Pronoun References.

3-4-5-1. Generic use of a pronoun

The 'generic' symbol '[g]' is marked when 'we', 'you' or 'they' are used to be more or less synonymous with 'one':

> [A020 10] "The only other thing >< **[g]** you can say about influenza is that it's totally unpredictable.

There is one instance of 3PP marked as a generic in the AT, *he* in [A039 59]:

> [A039 57] When (93 Joe Nelson 93) was a little boy back in Covington, Ky., <REF=93 he used to sketch pictures on the covers of <REF=93 his school books.
> [A039 58] <REF=93 He left <REF=93 his books behind, and forgot about the pictures, when <REF=93 he went to work for the railroad.
> [A039 59] Who worries about pictures or books when >< **[g] he**'s steering a huge, smoking iron horse down the tracks at 80 miles an hour?

3-4-5-2. Primary reference and secondary reference

For 1PPs and 2PPs, the AT coding scheme makes a distinction between a **Primary reference**, i.e. a reference to the addresser (speaker/writer) or the receiver (listener/reader) of the text, and **Secondary reference**, i.e. a reference to a character in a quoted speech.

A Primary reference is marked with '[P]' (capital letter 'P'):

> [A003 1] (2 Motorists 2) : concentrate on the car, but keep <REF=2 **[P]** your eye on the road.

A Secondary reference is marked with '[S]' (capital letter 'S'):

[A020 35] ">REF=27 **[S]** I'm already on the team," (27 Walcher 27) said confidently.

Since news reports are usually expressed in third person narration and most of the 1PPs and 2PPs appear within quoted speech, 1PPs and 2PPs in the AT are usually Secondary references. This distinction may be redundant where the quotation marks are present, but it may become necessary in particular in a Free Direct Speech construction which has no quotation marks, such as in the example below:

> He said I'll come back here to see you again tomorrow.
> (taken from Leech and Short, 1981:322)

Inclusive and exclusive of addressees in plural 1PP references

For a plural 1PP reference inclusive of the addressee, the symbol '[i]' is marked, and for exclusive of the addressee, the symbol '[x]' is marked. Notice that in the following examples, inclusive/exclusive feature markers are bundled with the Secondary character markers [S]:

> [A001 96] "<REF=103 **[Sx]** We are very grateful to <REF=104 **[Sp]** you."

> [A022 83] "<REF=92,97 **[Si]** Our common objective is to preserve peace and the spirit of liberty," he said of the U.S.-West German alliance.

Singular/Plural in 2PP references

For a singular 2PP, the symbol '[s]' ('s' is lower case) is marked:

> [A010 115] "Did <REF=98 **[Ss]** you ever take a bribe?" he was asked.

For a plural 2PP, the symbol '[p]' ('p' is lower case) is marked:

> [A056 3] U.S. District Judge Carl Rubin had told (7 jurors 7), "It is important that <REF=7 **[Sp]** your understanding of this case be limited to what is said in the courtroom."

Marking for Problematic cases

The symbol **[diff]** is marked on certain problematic pronouns.

The symbol '><' is used not for 'exophoric' or 'deictic' senses but simply used to mark an 'antecedentless' pronoun.

3-5. Alternative interpretation of the AT notations

Although the AT coding scheme is influenced by Halliday and Hasan's coding scheme (1976), since the AT is aimed to be theory neutral, the notations of the AT can be interpreted not only in terms of cohesion theory but also from a different theoretical perspective.

3-5-1. Identification of the Antecedent of a Pronoun

For a personal pronoun reference, the AT coding scheme assigns the three types of marker:

(i) Reference-direction marker: anaphoric '<' cataphoric '>'
(ii) Cohesive type marker: 'REF'
(iii) Index number: e.g. '109' in "<REF=109 her"

A reference-direction marker (anaphoric'<' or cataphoric '>') tells us that the antecedent NP can be found in the preceding text ('<') or the following text ('>'), and an index number tells us that the pronoun is coreferential with the NPs marked with the same index number.

Yet the three markers do not specify which particular NP is the antecedent of the pronoun. In a sense, the identification of the antecedent is left to each user's decision. For instance, consider the following four sentences at the beginning of a text:

[A003 98] Tennis star (109 Chris Evert Lloyd 109) has begun <REF=109 her vacation earlier than planned.
[A003 99] Illness forced <REF=109 her from a tournament Saturday night.
[A003 100] (109 Lloyd 109) had said <REF=109 she wanted to play through the finish of the $150.000 tournament Sunday and then begin an extended vacation from professional tennis.

[A003 101] **<REF=109** She has complained <REF=109 she lacks enthusiasm for the game.

We can identify the antecedent of the 3PP *She* in the last sentence [A003 101] from different points of view (cf. 2-10-2) as below:

[1] We can identify the antecedent to be the **nearest coreferential non-pronominal NP** (in this case the NP *Lloyd* in the previous sentence).

[2] We can identify the antecedent in terms of **regular coreferential relations**. In this case, based on the syntactic parallelism, we can identify the antecedent of the 3PP *She*, which is the main subject, to be the NP *Lloyd*, that is the main subject of the previous sentence.

[3] We can identify the antecedent in terms of **initial versus subsequent mention**. In this case, the antecedent NP of the 3PP *She* is the initial mention NP *Chris Evert Lloyd* in the first sentence [A003 98].

3-5-2. A cognitive interpretation of reference-direction marker

Rather than regarding the reference-direction marker as an indicator to point to a specific antecedent NP, we can interpret the reference-direction marker in a more cognitive manner. For instance, we can interpret the anaphoric marker ('<') in accord with the definition of anaphora stated by Cornish:

> Anaphora ... is a signal to continue the existing attention focus already established.
> (Cornish, 1996:22)

In the example above ([A003 98] to [A003 101]), as there is no other competing

referent candidate for female 3PPs, there is no discontinuity of the female-focus (highlight) on 'Chris Evert Lloyd'. Even if the main subject of [A003 100] *Lloyd* is replaced with "She", it does not affect the identification of the referent. We can say that 'Chris Evert Lloyd' is salient (highlighted) in a reader's consciousness throughout the discourse fragment.

In this respect, we can interpret the anaphoric marker '<' assigned to the 3PP *She* in [A003 101] to be indicating the continuation of highlight on the previously focused 'Chris Evert Lloyd', rather than to be identified with any particular NP. I will come back to this point when I discuss the borderline cases in terms of reference-direction judgement in chapter 7.

3-5-3. Re-interpreting index numbers to be 'addresses' of mental representations

In the AT, all the coreferential NPs are assigned identical index numbers, which form a coreferential chain of NPs (a 'cohesive tie' in Halliday and Hasan's term). From a cognitive point of view, it is possible to interpret an index number as the 'address' of the mental representation (MR) of a discourse entity which is the referent of the coreferential chain of NPs. (The discourse entity is a semantic abstraction, not a syntactic one: see 7-1, Halliday and Hasan, 1976:314) In the above example, we can interpret the index number '109' as the 'address' of the mental representation of 'Chris Evert Lloyd', to which all the NPs in the coreferential relation are linked.

3-6. The AT Annotations as a Record of 'Off-Line' Analysis.

In the discussion of the corpus based approach (cf. 1-2-1), I emphasised that we need a record of the interpretation of a text for the scientific investigation of text interpretation, since we cannot directly observe the process and the results of how people interpret texts.

I also stressed that the discourse annotation of the AT, where the analytic judgments of native speakers are recorded through the annotations, is one type of record of the interpretation of texts. In this section, I will discuss further the nature of the record of interpretation provided by the AT in terms of the process and the result of interpretation.

It is necessary to make clear the distinction between the **process** of interpretation and the **result** of interpretation. In the discussion of the corpus based approach (1-2-1), following Leech (1991:108), we distinguish between the process of writing (psychological process) and the product of writing (texts). In the same token, we can distinguish between the **process of interpretation** of texts and the **result of interpretation** of texts. We can investigate the issues of text interpretation focusing on the **process** of interpretation and the **results** of interpretation.

The process of interpretation (reading), as well as the process of production (writing), involves dynamic psychological aspects. The analysis of the process of interpretation concerns how a text is interpreted by a human processor as the text unfolds (on-line). The **on-line** text analysis concerns the sequential linear processing

of texts by a human processor.

The analysis of the **result** of interpretation of texts, on the other hand, is a static, ex-post-facto approach. After we finish interpreting a text and look back at it again, then we can identify which pronoun is coreferential with which full NPs in the text. In the terms of cohesion theory, we can identify cohesive ties in the text retrospectively, after interpreting it. This ex-post-facto approach can be said to be an **off-line** text analysis, in contrast with the **on-line text** analysis. The text analysis put forward by Halliday and Hasan is an off-line analysis which concerns not the process but the **result** of text interpretation.

What the AT provides is not a record of on-line analysis but a record of **off-line** analysis.

To observe the difference between the record of on-line analysis and that of off-line analysis, consider:

[A054 4] A Western diplomat who picked up an injured man in the street and took
him to Joumouriet Hospital said he ...

At the point when an analyst encounters the 3PP *he*, it may be difficult for the analyst to decide whether the referent is 'A Western diplomat', who is mentioned as the main subject, or 'an injured man', who is most recently mentioned by a pronoun (*him*). The record of the analysis of the referent of the 3PP may be either 'a Western diplomat' or 'an injured man' at this point.

After processing the rest of the sentence:

[A054 4] A Western diplomat who picked up an injured man in the street and took him to Joumouriet Hospital said he counted more than 50 bodies lined up on the floor for burial.

The analyst must now be quite sure the referent of the 3PP *he* has to be 'Western diplomat', hence the record of the analysis of the 3PP *he* is no longer ambiguous at the end of the sentence. If the judging points are marked as **(a)** and **(ß),** as in [A054 4'] below:

[A054 4'] A Western diplomat who picked up an injured man in the street and took him to Joumouriet Hospital said he **(a)** counted more than 50 bodies lined up on the floor for burial. **(ß)**

the record of the analysis of the referent of the 3PP *he* for each point would be:

Point (a):

referent = 'Western diplomat' (e.g. with about 60% probability)

referent = 'an injured man' (e.g. with about 40% probability)

Point (ß):

referent = 'Western diplomat' (e.g. with about 99% probability)

While the on-line approach concerns both points, the off-line approach concerns Point (ß) only.

We can see from this example that the interpretation of a pronoun depends on the point of judgement at which the interpretation is made. On-line judgement can

vary according to the judging point. Hence if we record (annotate) the analysis of the referent of a pronoun on an on-line basis, we need to encode different annotations for each judging point; which results in not a single annotation but a multiple annotation for a pronoun reference. This means that the single annotation, such as the AT annotation, is made not on an on-line basis but based on the final judgement, which is made towards the end of a sentence. In this respect, an off-line analysis can be considered as a special type of on-line analysis which focuses on the final judgement only.

An off-line analysis often involves a second pass (or further passes) through the text and the correction of an analysis made in the first pass, as Garside illustrates:

> ... analysts turn out often to want to do a second pass through the text to make corrections, after thinking "off-line" about particular annotation difficulties. ... The analyst makes a first pass over the whole text, marking the cohesive structures found there. There will usually be difficulties in deciding how to mark certain features. The analyst can re-edit the partially-marked text in due course ...
> (Garside, 1993)

On-line approaches, on the other hand, usually concern only the first pass over a text. If difficulties of interpretation arise for a reader during the first pass, on-line approaches try to analyse the conditions which trigger the difficulties.

It should be stressed here that these two approaches are not incompatible but need to be complementary.

Recently some cognitive discourse theorists (Brown and Yule 1983, Emmott 1989,

1997; Cornish 1986b, 1996, Stockwell 1995) have criticised the theoretical concepts used in off-line analysis as in Halliday and Hasan (1976).

Although some of these concepts are inadequate in terms of human text processing and some of them cannot be applied to on-line analyses, I would argue that there is no reason to disregard the off-line analysis altogether. For instance, regular coreferential patterns, such as the initial adverbial cataphora patterns (see 4-1), can be regarded as part of a reader's interpretative knowledge, based on which a reader can make a prediction of what is likely to occur in the unfolding text (on-line). Hence such regular coreferential patterns cannot be ignored in on-line analyses.[14] Unless we complete the interpretation of a text or a portion of the text, it may be difficult to identify the coreferential or co-specifying relations within it.

This is one of the tasks an **off-line** text analysis can undertake.

Hence I would argue that on-line text analyses need to be complemented by off-line text analyses.

---

[14] Another argument in favour of both on-line and off-line analysis, pointed out by G. N. Leech (in an informal conversation), is as follows: There is a basic difference between the processing of speech and of a written text. In interpreting a written text, a reader always has the opportunity to re-read a passage to overcome any problems of interpretation. In this sense, a reader (but not a speaker) can process anaphoric relations both on-line and off-line.

3-7. Corpus Data retrievable from the AT

From the Anaphoric Treebank, using appropriate computer programs, we can more or less automatically obtain corpus data which can be of use to the study of discourse anaphora and automatic pronoun resolution.

Tables illustrating how discourse entities are mentioned as the discourse unfolds:

e.g.

[Note: Asterisks indicates that the NP occurs in direct speech]

Sentence	Mentions of 'Mark Twain'	Mentions of 'William Pizzano'	Mentions of 'Massachusetts Public Works Department'	Mentions of 'Joe Santa Fe'
A014 39	Mark Twain			
A014 39	he			
A014 44		William Pizzano		
A014 44			the Massachusetts Public Works Department	
A014 44		he		
A014 44			* We	
A014 45			* We	
A014 46			* We	
A014 46			* we	
A014 47		his		
A014 47			his department	
A014 48			it	
A014 53				* I
A014 53				Joe Santa Fe
A014 54				* I

List of referent candidates for a pronoun

Since the Anaphoric Treebank is syntactically parsed in the form of labelled bracketing, any kind of phrases such as noun phrases or verb phrases can be picked up

automatically. This enables us to produce a list of referent candidates for a pronoun automatically.

Identification of the head of a noun phrase:

The parsed texts enable us to identify, to a large extent, the head of a noun phrase, which can be useful in checking the Number agreement between a pronoun and its antecedent candidates.

Data of distance between anaphors and antecedents

The Anaphoric Treebank also enables us to count the number of intervening sentences, noun phrases, and words between anaphors and antecedents.

3-8 Size of the Anaphoric Treebank (AT) used in this thesis

The Anaphoric Treebank used in this thesis consists of sentences from [A001 1] to [A060 89].

Total number of texts:	481
Total number of sentences:	6418
Total number of words:	129,218
Total number of personal pronoun:	4809
anaphorically marked personal pronoun	4666 (97.0%)
cataphorically marked personal pronoun	133 (2.8%)
personal pronoun marked as indeterminate (anaphoric or cataphoric)	10 (0.2%)

3-9 Conclusion

It is inadequate to account for pronoun reference solely in terms of a sentential (syntactic) approach. What is needed is a study of naturally occurring text data. The issues related to the interpretation of texts also need a record of interpretation. Creating a corpus with discourse annotation, such as The Anaphoric Treebank, is an attempt to meet these needs. Fligelstone, a member of the staff of the AT project, notes:

> ... the information we are encoding offers considerable benefits to those researching in the field of anaphora who want to base their studies and experimentation on observed data.
> (Fligelstone, 1992)

One of the main aims of this thesis is to demonstrate the usefulness of the corpus with (off-line) discourse annotation for the empirical inquiry of pronoun reference.

In the next section, I will present the regular coreferential patterns observed in the cataphorically marked pronoun references in the AT.

Chapter 4: Corpus data of cataphora (backwards anaphora): Structural patterns of occurrences

4-0. Introduction

Although cataphora has been discussed mainly on the basis of invented, isolated single-sentence examples, there are a few cases where it has been investigated by collecting naturally occurring corpus data. In this section, I will look at the corpus data of cataphora as it has appeared in these exceptional works, as well as the data of cataphora retrieved from the Anaphoric Treebank (AT), trying to identify regular structural patterns observed in them. 4-1 includes a brief overview of data presented in Carden (1982), Kanzaki (1997), and Van Hoek (1997). 4-2 presents the cataphora data extracted from the AT. Those data are compared and discussed in 4-3.

4-1. Corpus data of cataphora presented in other research reports.

4-1-1. Backwards anaphora corpus data presented by Carden (1982)

Carden (1982) gives an account of cataphora based on a significant amount of naturally occurring samples (over 800 examples). The main purpose of his collection of cataphora data is to provide conclusive counter-evidence against what he calls the 'Forward Only Hypothesis'[15] that claims that the referent of seeming cataphoric pronouns must have been mentioned in the preceding discourse, and hence must be predictable. To this end, Carden tries to present cases of so-called **genuine backwards anaphora**, "where the backwards pronoun in principle could not have had a referent earlier in the discourse"(Carden, 1982). His data include mainly those cases where the referent of a cataphoric pronoun is the first-mentioned entity in the discourse (First-mention).

Carden gives five criteria for **genuine backwards anaphora**, each of which is listed with an example below.

[Note: Cataphoric pronouns and their antecedents are italicised in the examples]

First Mention: The antecedent is a definite NP   (over 100 instances)

(8) To the Editor:
   After *his* recent election as Republican national chairman, ***Bill Brock*** said ...
   (Letter, *New York Times* (NYT), 28 Jan. 77)

---

[15] 'Predictability Requirement' in Kuno(1972, 1975)'s term.

85

<u>First Mention: The antecedent is an indefinite NP (29 cases)</u>

(12) When *she* was five years old, *a child of my acquaintance* announced a theory
that she was inhabited by rabbits.
(*New York Times*, 6 Nov. 78)

<u>Singular Pronouns Bound by Singular Quantifiers[16] (11 cases)</u>

(17) . . . there are six legally operated and licensed poker cardrooms . . . As *its*
major source of income, *each club* collects a playing fee from the
players every half hour. . .
(*Social Problems*, 28: 557 (1977))

<u>Plural Pronouns with the Variable Reading[17] (17 cases)</u>

(21) Did you know that when *their* wives leave *them*, *two men in five* go
bananas?
Lanniglan's Rabbi, TV, fall 1977
(collected by T. Dieterich)

<u>Variable-Reading Pronouns Bound by Generics (30 cases)</u>

(22) No matter how innocent *he* may be in *his* inner soul and in *his*
motivations, *the effective mathematician* is likely to be a powerful
factor in changing the face of society.
(Weiner, *Ex-Prodigy*: 189-190)

---

[16] With regard to a singular pronoun whose antecedent is modified by singular
quantifiers such as *every, each, any* or *none*, Carden claims that the pronoun
represents a variable bound by the quantifier, and "in principle its antecedent must lie -
in the quantifier phrase, whether or not the set quantified over appears in the
preceding discourse(p.369)".

[17] Carden claims that a plural pronoun can also be interpreted as a variable bound by
the quantifier in certain context. (1982:369)

Carden(1982) presents his data not according to the regular structural patterns of cataphora but according to the criteria for the genuine backwards anaphora. To obtain, at least, a rough estimation of regular structural patterns of cataphora, I analyse the 20 representative samples which appeared in his article[18] in terms of the grammatical position of the pronoun-antecedent pattern, summarised in the following table 4.1.1.

Table 4.1.1: Carden (1982)'s 20 samples in terms of the grammatical position of the pronoun-antecedent pattern.

Pronoun position	Antecedent position	freq.
Initial Adverbial Clause	Subject of the main clause (MC)	8
Initial Prepositional phrase	Subject of MC	5
other patterns		7
Total		20

As shown in table 4.1.1, 13 out of 20 (65%) occurrences of cataphoric pronoun reference involve the pronouns occurring in initial adverbials and antecedents occurring as the main subject. This pattern (henceforth initial-ADV type) is the typical construction of what he calls the 'structurally-governed type' (1982:379).

In addition to the over 187 first mention cases, Carden reports 45 occurrences of what he calls 'S-Pronominalization' pattern, as below.

(32) Then Griddlebone she gave a screech, for she was badly skeered; I am
    sorry to admit *it*, but *she quickly disappeared*.
    (Eliot, 'Growltiger's Last Stand',

---

[18] Each example is listed with its source in Appendix 4-1-1.

87

4-1-2. Corpus data of cataphora presented by Kanzaki(1997).

Along the same lines of Carden (1982), Kanzaki(1997) presented 24

cataphoric pronoun examples in which the referent of pronoun is not predictable from

the preceding contexts. Most of them occur at the beginning of texts or paragraphs.[19]

Table 4.1.2 summarises his examples according to the grammatical position of the

pronoun-antecedent pattern.

Table 4.1.2: Kanzaki (1997)'s examples in terms of the grammatical position of the pronoun-antecedent pattern

Pronoun position	Antecedent position	freq.
Initial Adverbial Clause	Subject of the main clause (MC)	12
Initial Prepositional phrase	Subject of MC	9
Genitive of Subject of MC	Object of MC	1
Initial Adverbial Clause	Genitive of Subject of MC	1
Initial Adverbial Clause	Object of MC	1
Total		24

As shown in 4.1.2, 88% of Kanzaki's examples are of the initial-ADV type,

exhibiting the same tendency of occurrence as Carden's samples.

---

[19] Kanzaki(1997)'s samples are collected from newspapers, magazines, academic writings and novels.
Each sample is listed with the source of reference in Appendix 4-1-2.

4-1-3. Backwards anaphora corpus presented by Van Hoek (1997:112-114)

Van Hoek (1997:112-114) presents a survey of 500 examples[20] of **genuine**

cataphora collected according to the same criteria as Carden (1982). Van Hoek says:

> The criterion for inclusion in the corpus was that there must be no plausible
> antecedent for the pronoun other than the noun which appeared to its right.
> (Van Hoek, 1997:109)

Her objectives of collecting cataphora data, however, is not only to refute the

'Forward Only Hypothesis' (Carden), but also to give an account of cataphora in

terms of the reference point model of the cognitive grammar developed by

Lagnacker(1987, 1991). One of her main arguments is that "backwards anaphora

typically involves a significant asymmetry in prominence between the nominal

conceptions corresponding to the pronoun and the antecedent"(1997:108). To this

end, Van Hoek presents her data according to the structural pattern of the pronoun-

antecedent occurrences, as summarised in Table 4.1.3.a.

Table 4.1.3.a: Summary of Van Hoek (1997)'s cataphora data according to the structural pattern of the pronoun-antecedent occurrences

	Pronoun position	Antecedent position		
1	possessive pronoun in initial adverbial phrase	subject of Main Clause (MC)	307	61 %
2	subjective pronoun in initial adverbial clause	subject of MC	95	19 %
3	Other position	Other position	98	20 %
Total			500	100 %

---

[20] Van Hoek says:
"Examples were collected from almost every imaginable source: magazines, novels, newspapers, placards at museums, sign on buses, etc." (Van Hoek, 1997:109)

Table 4.1.3.a exhibits the predominance of initial-ADV type (80%) in her data, the

same tendency of occurrences observed in Carden (4.1.1) and Kanzaki (4.1.2).

According to Van Hoek (1997), in Cognitive Grammar (CG), specific grammatical

patterns are captured "in terms of constructional schemas, templates which speakers

extract from exposure to actually occurring expressions" (1997:111). Regarding the

predominance of the initial-ADV type in her data as well as Carden and Kanzaki's

data, this cataphora pattern can be said to be "the most strongly entrenched schema

for backwards anaphora" (Van Hoek, 1997:113).

In addition to the pronoun-antecedent pattern, Van Hoek provides frequency

data of the grammatical position of antecedent NPs, summarised in Table 4.1.3.b.

Table 4.1.3.b: Grammatical position of antecedent NPs in cataphora data (Van Hoek, 1997).

	Antecedent position		
1	Subject of the main clause	423	84.6 %
2	Complement (direct object) of main verb of either main clause or subordinate clause	58	11.6 %
3	The object of an of-phrase attached to one of the arguments of the verb	7	1.4 %
4	The possessor attached to the subject nominal	5	1.0 %
5	Other position	7	1.4 %
	Total	500	100 %

This data clearly shows that the subject position is preferred as the antecedent

position in cataphoric reference.

In addition to First-mention cases, Carden(1982), Kanzaki (1997), and Van

Hoek (1997) also present a few cataphora examples where a pronoun is used in

reference to a previously mentioned discourse entity (henceforth 'Previously-

mentioned' type). The previously-mentioned cases presented by them involve cases of pronoun reference in which a cataphoric reading resolves possible ambiguity of reference, or overrides a reading with reference to competing referent candidates. Consider the following example taken from Van Hoek (1997:110):

> 1a. And **Naiibullah** never misses a chance to manifest his piety in public. In *his* letter to Gorbachev, ***Khomeini*** referred to the failure of the atheist state in the Soviet Union to destroy Islam. (The Nation 10/ 16/89)

In (1a), even though **Khomeini** was mentioned in the previous discourse, the previously established focus at the point of encountering the 3PP *his* is not **Khomeini** but **Naiibullah**, which is mentioned as the subject of the previous sentence. In this 3PP reference (*his*), the cataphoric reading linked with the following main subject *Khomeini* overrides the reading with previous-focus on **Naiibullah**. In another example (1c) taken from Van Hoek (1997:110):

> 1c. **Colie** is a seeker. **Robin** is a discoverer. With a major college scholarship awaiting *her* signature, ***Colie*** has grabbed the brass ring. (San Diego University City Light 2/89; Robin is also female)

Colie and **Robin,** can both be the candidate referent of 3PP *her* (in *her signature*), which causes ambiguity of the reference. In this case, the cataphoric reading with the following main subject *Colie* resolves the ambiguity. Like First-mention cases, these Previously-mentioned cases where the cataphoric reading has an overriding effect (henceforth 'Overriding type') can also be identified as cases of **genuine backwards anaphora.**

4-2. Corpus data of cataphora retrieved from the Anaphoric Treebank

4-2-0. Introduction

Before looking at the cataphora data retrieved from the AT, attention should be paid to the difference in the nature of data between the AT data and the data presented in 4-1 by Carden(1982), Kanzaki (1997), and Van Hoek (1997).

(1) *Source of data.*

The source of the Anaphoric Treebank (AT) is the American associate press corpus, which includes news reportage texts only. In contrast, the data presented in 4-1 are taken from variety of genres of written texts, including not only news reportage but also novels, magazines, advertisements etc.

(2) *Whole text or fragment of text.*

While the AT is a collection of short but complete texts[21], the Data presented in 4-1 are mostly collections of text fragments. This is partly because most of their data are First-mention cases; hence they do not need to include the preceding texts to demonstrate that these examples cannot be judged as anaphoric.

(3) *Annotated or not annotated.*

The main difference between the AT data and the data presented in 4-1 is that while the former are annotated, the latter are not. The AT data consists of the examples of pronoun reference judged as cataphoric by the analysts of the AT. The discourse annotation of the Anaphoric Treebank (AT) can be seen as a record of interpretation; it is a method of recording how the text is interpreted by the analysts of the AT. In this respect, the AT includes a kind of record of native speaker intuition similar to informant tests. In other words, the AT includes a kind of psychological data as well

---

[21] Some parts are missing in a few texts of the AT.

as textual data. As discussed in 1-2-1, we can never know how a text is interpreted by other people simply because we cannot directly access and observe other people's mind or intuition. We need, therefore, some record of responses to the text made by readers in order to investigate the interpretation of text. As the judgement of the direction of pronoun reference (anaphoric/cataphoric) is part of the interpretation of texts, we need the records of such interpretation of texts to investigate how the anaphora/cataphora judgements are usually made. Corpus data with discourse annotation is provided by the Anaphoric Treebank (AT). Creating the AT can be seen as an attempt to meet such needs.

This methodological characteristic of the Anaphoric Treebank (AT) has a practical consequence. As mentioned before, the data presented in 4-1 by Carden(1982), Kanzaki (1997), and Van Hoek (1997) are limited to the least disputable examples (genuine backwards anaphora). They have to limit the data presented in their articles because they need to argue against prevailing scepticism towards cataphora manifested in 'Predictability requirements' (Kuno, 1975)[22]. This implies that they could have included data other than these least disputable genuine cataphora (First-mention and Overriding type). In other words, there are instances somewhat 'weaker', rather more arguable Previously-mentioned cataphora, which they avoid to include. Van Hoek (1997:129) notes the possibility of such 'weaker' Previously-mentioned cataphora, when she tries to distinguish between genuine cataphora and 'repeated identification' cases[23]:

---

[22] Or the 'Forward Only Hypothesis' in Carden(1982)'s term.
[23] 'Repeated Identification' cases involves a pronoun precedes a coreferential full NP within a sentence, but the referent of the pronoun is already mentioned in the preceding discourse and to a large extent predictable.

Without the capability of reading the author's mind, however, it was not possible to be sure of filtering out every example of genuine backwards anaphora. No doubt some of these examples were intended to be read as backwards anaphora, with the full noun phrase as antecedent, even though the pronoun could also have picked up its reference from the preceding discourse. (Van Hoek, 1997:129)

The cataphora data retrieved from the AT include not only genuine cataphora (First-mention and Overriding types) but any instance of pronoun reference which is judged as cataphoric by the AT analysts. For instance, consider the initial-ADV 3PP reference (*he*) in [A024 10] (judged as cataphoric by the AT analyst):

[A024 8] *Johnson* scored nine straight points early in the second half to stretch out a 54-36 Tennessee lead with 12:40 remaining.
[A024 9] The Bulldogs could get no closer than nine.
[A024 10] When *he* was benched with 10 minutes still left in the game, *Johnson* had a college career total of 1,982 points, 20 above King's three-year total.

It is possible to argue that this 3PP reference has to be anaphoric on the ground that there is no candidate referent for this 3PP *he* other than **Johnson** which is focused just before the previous sentence; hence although previous focus is not on **Johnson** but on **The Bulldogs, Johnson** can still be salient enough to be the referent of 3PP *he*. The 3PP *his* in the initial-ADV pattern in [A047 66] below, judged as cataphoric by the AT analyst, could be more arguable:

[A047 65] In 1966, *Lauck* retired to *his* Lake Hamilton home near Hot Springs.
[A047 66] Since *his* retirement, *Lauck* had served a five-year stint on the state Racing Commission .

In this case, not only is **Lauck** the only candidate referent for the 3PP *his*, but also **Lauck** is previously focused. For these instances, however, a counter-argument can be made as follows: if the writers do not intend cataphoric reading for these 3PP references, why do they choose full NPs as the main subject rather than pronouns ?

This point can be shown if the main subjects of these examples are replaced with appropriate pronouns, as in [A024 10]' and [A047 66]':

[A024 8] *Johnson* scored nine straight points early in the second half to stretch out a 54-36 Tennessee lead with 12:40 remaining.
[A024 9] The Bulldogs could get no closer than nine.
[A024 10]' When *he* was benched with 10 minutes still left in the game, *he* had a college career total of 1,982 points, 20 above King's three-year total.

[A047 65] In 1966, *Lauck* retired to *his* Lake Hamilton home near Hot Springs.
[A047 66]' Since *his* retirement, *he* had served a five-year stint on the state Racing Commission .

Since the same pronoun references can be obtained from the replaced versions as in the original examples, the writers could have used pronouns as the main subject instead of full NPs if they had not intended the cataphoric reading. It may be the case, therefore, that the typical initial-ADV cataphora pattern influences the choice of referring expressions by writers as well as the interpretative strategy of the readers. This thesis does not deal with the issues related to the writer's perspective such as the choice of referring expressions. In any case, the cataphorically judged pronoun references in the AT seem to range between the least disputable cases of genuine cataphora and the arguable cases shown in [A047 66]. The cataphora data retrieved from the AT include not only cases of genuine cataphora but also other Previous-mention cases that Carden(1982), Kanzaki (1997), and Van Hoek (1997) did not include in the data presented in their articles. It can be said that the AT cataphora data provide psychological evidence that cataphoric reading is possible not only for genuine cataphora cases but also for those 'weaker' arguable cases.

4-2-1. First-mention cataphora in the AT

As well as the data presented in 4-1, the cataphora data retrieved from the AT include 47 First-mention cataphoric pronouns, out of 133 total occurrences of cataphora (35.3%). (47 First-mention cataphora cases are listed in Appendix 4-2-1, and 86 Previously-mentioned cases are listed in Appendix 4-2-2.)

First-mention cataphora involving 3PP:

[A047 91] For *his* part , **American Coach Herb Brooks** has his own opinions on the Soviet style and its practicality for North American players.

First-mention cataphora involving singular 1PP:

[A015 42] "*I* heard him yell, Where's grandma, where's grandma ? " ' said *his mother, Jacqueline Hodge , 25.*

Among the 47 cases of First-mention cataphora, 6 instances are appeared at the beginning of the texts:

[A011 57] Predicting *he* will fare " much better than anyone expects " in Sunday's Maine caucuses, **California Gov. Edmund G. Brown Jr.** Tuesday compared his Democratic presidential rivals to " two peas in a pod . "

Personal pronouns have been described as linked typically with information that is given or recoverable, as stated by Palmer:

> We can avoid restating in detail what is given by using pronouns — the third person pronouns *he/she/it/they* instead of the already mentioned *the little boy, the man on the corner*, etc.   (Palmer, 1981:160)

The reference of a person pronoun is usually assumed to be "made to something

which is given or known within the linguistic or situational context" (*CGEL*:335).

In writing, the reference of 3PP is usually assumed to be made to something already mentioned in the preceding text either directly or indirectly.

The corpus data of cataphora overviewed in 4-1 and extracted from the AT, however, clearly show that personal pronouns can be used for reference not only to given/recoverable information but also to new information (in particular First-mention cases) in written discourse. In other words, personal pronouns can be used to introduce a new entity in the discourse, and in this connection, Van Hoek makes the following suggestion:

> I suggest that one of the functions of backwards anaphora may be the concise introduction of new information into a discourse (here meaning written text), and that typical backwards anaphora constructions serve to make a referent somewhat more accessible (in Chafe's sense) before it is fully introduced with a full noun phrase. (Van Hoek, 1997:114)

First-mention cataphora also has a theoretical impact upon those cognitive discourse theories that assume that pronoun reference is made to already-activated (or semi-activated) mental representations in the memory, since no such mental representation is available for First-mention cataphora. I will discuss this point in chapter 6.

4-2-2. Cataphorically marked cases in the AT according to the structural pattern of the pronoun-antecedent occurrences.

Table (4.2.2) cataphorically marked cases in the AT according to the frequent structural pattern of the pronoun-antecedent[24] occurrences.

Table (4.2.2) Cataphorically marked cases in the AT according to the frequent structural pattern of the pronoun-antecedent occurrences

	Pronoun position	*Antecedent position*	freq.	%
1	Pronoun position: Singular 1PP in the initial direct speech	Antecedent position: Subject of reporting clause	59	45 %
	e.g. [A022 4] "*I* knew John DeLury," said *Mayor Edward I. Koch* on learning of his death.			
2	Pronoun position: 3PP in the initial adverbial (initial-ADV)	Antecedent position: Main subject	46	34 %
	e.g. [A026 41] In *his* news conference Wednesday, *President Carter* dismissed a suggestion raised Monday ...			
3	Pronoun position: Plural 1PP in the initial direct speech	Antecedent position: Oblique position in the reporting clause	16	11 %
	e.g. [A029 43] "I thought *we* had the game won in regulation," said *North Carolina* Coach Dean Smith.			
4	Pronoun position: Other position	Antecedent position: Other position	12	10 %
	Examples are listed in 4-3.			
	Total		133	100 %

The most frequent pattern of cataphora (45%) in the AT involves a first person singular pronoun (1PSP) in the initial direct speech (DS) position with its antecedent as the subject of the reporting clause (henceforth '1PSP-initial-DS' pattern). This is the major difference, in the occurrences of structural patterns of cataphora, between

---

[24] I assume that the 'antecedent' of a cataphorically marked pronoun is the first coreferential full NP appearing to the right of the pronoun.

the AT data and the data presented in 4-1. There is no instance of the 1PSP-initial-DS pattern in the data reported in Carden (1982), Kanzaki (1997), and Van Hoek (1997). The initial-ADV pattern, the most predominant pattern observed in 4-1, is the second most frequent (34%) in the AT data. Both predominant patterns (1PSP-initial-DS and initial-ADV) involve the antecedent at the subject position of the main or reporting clause. The preference of the subject as the antecedent position in cataphora (79%) can be seen in the AT data as well as in the data presented by Van Hoek (1997).

The 1PSP-initial-DS cataphora pattern may reflect the straightforward linguistic knowledge that the addresser of a piece of direct speech (DS) is usually mentioned in the subject of the reporting clause, as well as in singular 1PPs within DS. The singular 1PP (1PSP) within DS usually corefers with the subject of the reporting clause.

In news reportage, a piece of direct speech often precedes the reporting clause. When a piece of DS precedes the reporting clause, the interpretation of a 1PSP within DS is more or less suspended until the subject of reporting clause is identified. There is no guarantee that initial DS is always followed by a reporting clause. When there is no reporting clause, the 1PSP reference within DS has to be resolved by inference from the preceding context.

A notable difference between the 1PSP-initial-DS pattern and the initial-ADV pattern is that the former involves not only intra-sentential references but also cross(inter)-sentential references.

Consider the 1PP *I* in [A041 25] below:

[A041 22] *Malmquist*'s final point total was 391.915.

[A041 23] Although **Wehling** only managed a ninth-place finish in the 15-km race, it was enough to give **him his** third gold medal in the Nordic Combined event.

[A041 24] **His** point total was 432.200.

[A041 25] " *I* wish *I* could be like Bill Koch.

[A041 26] He has so much intestinal fortitude, " said *Malmquist*, referring to the top U.S. cross-country skier.

The 1PSPs in [A041 25] are analysed as cataphoric (linked with *Malmquist* in [A041 26]) by the analyst of the AT. Although their referent **Malmquist** is already mentioned in [A041 22], these 1PP references are unlikely to be anaphoric because the previously focused (highlighted) item, which is usually assumed to be the antecedent of an anaphoric pronoun, is not *Malmquist* but *Wehling*, who is referred to by the pronouns three times in the previous two sentences ([A041 23] and [A041 24]). This 1PP reference in [A041 25] is a clear case of cataphora overriding a previously focused competing referent candidate (Overriding type), and it is not intra-sentential but cross-sentential. The same point can be made about the 1PSP *I* in [A030 7] below:

[A030 5] **Ramirez** and *Lara* arrived last Wednesday from Palm Springs, Calif., where they played in a Grand Prix tournament.

[A030 6] **Ramirez** withdrew from that tournament after a slight back injury caused by a distended ligament but was reported in satisfactory condition by the team's physician.

[A030 7] "*I* think we can win.

[A030 8] It is going to be very difficult but *I* think we can make it," said *Lara*.

Both *Ramirez* and *Lara* are candidate referents for the 1PSP *I* in [A030 7] and [A030 8]. Regarding the fact that the previous focus is not *Lara* but *Ramirez*, readers' prediction of the referent is more likely to be *Ramirez* than *Lara*. In this case, at least, readers may not be sure which candidate is the speaker of the quotation until

they encounter the subject of the reporting clause *Lara* at the end of [A030 8]. The

1PP *We* in [A038 71] below provides another example of cross-sentential cataphora:

[A038 65] In fact, the circus comprises only a few of the thousands of models
        Wagenfuehr has built - and still builds.
....
[A038 70] Once he built a tiny replica of *his* church in New Braunfels, complete with
        stone walls.
[A038 71] "*We*'ll never grow old this way.
[A038 72] *We*'ll never grow up," said *Wagenfuehr's patient wife of 18 years,
        Eleanor*.

The 1PP *We* in [A038 71] is coreferential with *Wagenfuehr* (which is mentioned six

sentences before) and *Wagenfuehr's patient wife of 18 years, Eleanor* in the next

sentence. Since the latter is firstly mentioned in the text, the reference of *We* in

[A038 71] is not predictable from the preceding text; hence it cannot be anaphoric but

cataphoric, and it is a cross-sentential reference.

These actual occurrences of cross-sentential cataphora seem to contradict the

claim made by Carden (1982) and Van Hoek (1997) that "ordinary backwards

anaphora includes both pronoun and antecedent within a single sentence" (Van Hoek,

1997:110).

One way to account for the initial DS cross-sentential cataphora is to identify

the piece of DS as subordinate to the reporting clause no matter how many sentences

the piece of DS includes. CGEL (pp.1022-23) provides an account of identifying a

piece of DS to be a subordinate clause functioning as a direct object of the reporting

verb. According to this analysis, an initial DS is a fronted direct object, often causing

the subject-verb inversion when the subject NP is 'heavier' than a pronoun

(CGEL:1380), as in:

[A005 53] "*I* think this contract is a major step forward, " said *Paul Silas, president of the Players Association.*

The inversion, however, does not necessarily occur, as below.

[A038 34] "Maybe it's *me, "an embittered Coach Bobby Kromm* said .

To identify a piece of DS as a subordinate clause may accord with the claim of intra-sententiality of cataphora made by Carden (1982) and Van Hoek (1997), in that both initial-ADV and 1PSP-initial-DS patterns can be identified as subordination.

It should be noted, however, that CGEL (p.1023) also provide an opposite view in which the reporting clause is identified as "subordinate, functioning as an adverbial" in particular when it occurs at the medial position or when it is omitted. For example,

[A009 38] "I'm enjoying all the affairs," **he said**, "but they 're coming to an end. [A009 39] It's time to get out into the real world."

If both direct speech (DS) occurring before and after the reporting clause (*he said*) are to be analysed as subordination, it would be difficult to represent the subordination involving both pieces of direct speech in a tree diagram.

It is even more problematic for the generative right-branching X-bar representation. (see Fig. 53.2.3 in 5-3-2-1) In the X-bar representation, an element occurring at the initial (preposed) position occupies either COMP, which is immediately dominated by S' (=C'), or SPEC, which is immediately dominated by S''

(=CP). The initial direct speech (DS) must be positioned at either COMP or SPEC, which are upper-left to the main clause and cannot be dominated or c-commanded by the reporting verb (said). In this respect, rather than analysing initial-DS as subordination, it would be more plausible to analyse the reporting clause (*he said*) as insertion. Also [A002 58] below cannot be accounted for as syntactic subordination since it does not have a reporting clause.

[A002 57] "I'm nobody," he said, when asked his name.
[A002 58] "I'm just here to see that Fess goes home."
[A002 59] Fess, or the Fessor, was born Henry Roeland Byrd.

Alternatively, this cataphora pattern can be accounted for in terms of the deictic condition, no matter whether the reference is intra-sentential or cross-sentential. In 2-2, a distinction is made between **cross(inter)-deictic coreference**, i.e. a coreference between an NP in the narrative and an NP in a piece of DS, and **intra-deictic coreference**, i.e. a coreference between NPs within the narrative or within a piece of DS. Following this distinction, the coreference between 1PSPs in direct speech and the subject of the reporting clause can be described as **cross-deictic**, and the coreference involving initial-ADV pattern within the narration or within direct speech can be described as **intra-deictic.**

The singular 1PP reference usually remains constant, regardless of the sentence boundaries or focus (highlight) shift, unless deictic shift[25] takes place. This means that the referent of a singular 1PP is assumed to be usually predictable and accessible, and hence anaphoric within the narrative or within direct speech (DS). Intra-deictically, a 1PP reference is unlikely to be perceived as cataphoric. In the AT,

---

[25] See 2-2 for the explanation of deictic-shift.

there are 8 instances of the initial-ADV pattern where a singular 1PP occurs both in the initial adverbial and in the main subject position (such as [A009 67] below), but none of these singular 1PPs in the initial adverbials are analysed as cataphoric.

[A009 67] " Once *<REF=82 I* got over the blue line , *<REF=82 I* didn't see anyone to pass to .

The initial-ADV cataphora pattern, which mainly involves 3PPs, usually occurs in the narrative:

[A023 68] By offering the motion as *his* own, **Rep. Richard Bolling, D-Mo**, was able to cut off debate before Republicans were ready.

This pattern can also appear within direct speech, as below:

[A043 82] "In some of >REF=107 *their* new passenger cars, (107 *the manufacturers* 107) will be using all-new types of chassis structural components, and A. O. Smith expects to be an important supplier of these products as well," Parker said.

[A035 59] "Once last year , after *>REF=76 we* lugged all the equipment up , *(76 three of us 76)* just fell on the snow and looked down at the clouds below .

In these examples; although it is not possible to know whether the speakers of DS intend cataphoric reading for these pronouns, the analysts of the AT treat them as cataphoric. In these cases, the initial-ADV cataphora pattern is recognised within initial DS. In other words, it can be said that the initial-ADV cataphora pattern as an interpretative scheme functions **intra-deictically.**

4-3. Other cataphora constructions

So far, we have looked at the predominant cataphora types (initial-ADV and 1PSP-initial-DS patterns) involving the antecedent as the clause subject. In the AT data, there are other, less frequent cataphora patterns in which the antecedent full NP does not occur in the clause subject position. Some of them could be identified as **preferred constructions of cataphora**, in addition to the predominant initial-ADV and 1PSP-initial-DS patterns. This section presents these less frequent patterns appearing in the AT data as well as the data presented by Carden (1982), Kanzaki (1997) and Van Hoek (1997) in the following order:

4-3-1. Intra-deictic coreference

    (a) Pronouns occur in an initial adverbial and the antecedent occurs in the main clause other than in the subject position.

    (b) The initial clause including the cataphoric pronoun is conjoined with the following clause including the antecedent by the conjunction *but*

    (d) Parenthetical construction

    (e) The pronoun appears in the heavier subject and the antecedent occurs as the object of the main verb

    (f) Postponed identification

    (g) Reporter's supplying information within DS

4-3-2. Cross-deictic coreference

    (h) Plural 1PP in the initial DS has its antecedent in the reporting clause.

    (i) 3PP in the initial DS has its antecedent in the reporting clause.

    (j) Tag question

4-3-1. Intra-deictic coreference

(a) Pronouns occur in an initial adverbial and the antecedent occurs in the main clause other than in the subject position.

Even if the antecedent position is not the subject of the main clause, the pronouns in the initial adverbial sometimes can still be recognised as cataphoric when the antecedent appears somewhere else in the main clause. The antecedent position in this pattern varies and it is difficult to identify the regular pattern in its occurrence.[26]

> Despite *its* long reach, the testing activities of *ETS* are not overseen by any public agency.
> (*Readers Digest* 3/80) — taken from Van Hoek (1997:116)

> By matching *his* parents' characters and backgrounds with the findings of scholarly studies, they decided that *Michael Lee Hall* faced a precise 40% chance of being abused within the next five years.
> (*Los Angeles Times* 2/1/89)   — taken from Van Hoek (1997:119)

In the following AT examples, the initial adverbials including the pronouns and the main clauses including the antecedents occur within a piece of DS.

[A026 10] On a trip to Japan to try to persuade the makers of Toyota and Datsun to build cars in the United States, Fraser told the Detroit News by telephone , " Given the relationship *they* already have, I think it would be better for *Mitsubishi and Chrysler* to have a joint venture."

[A017 42] "At *its* broadest point , I 'd say *the water* is about 1,000 feet wide .

---

[26] Further examples are included in Appendix 4-3-1.

This pattern often involves 3PP *it* as the anaphor and the entire following clause as the antecedent, a construction called 'S-Pronominalization' by Carden (1982). Carden (1982:372) reports 45 occurrences of 'S-Pronominalization', as below.

> Then Griddlebone she gave a screech, for she was badly skeered; I am sorry to admit *it*, but *she quickly disappeared*.
> (Eliot, *'Growltiger's Last Stand'*, — taken from Carden (1982:372)

Not only the entire clause but also the other elements can be the antecedent position of the pattern (b). For instance. Van Hoek (1997:124) presents what she calls the 'unequal conjoined structure' as below:

> *It* may not be great, but *'Footsteps'* isn't that bad.
> (*San Diego University City Light*, 2/2/89, title of article)
> — taken from Van Hoek (1997:124)

> *It*'s tabloid TV however you look at it, but at least *NSC'*s "Unsolved Mysteries" is striving to be different from "America's Most Wanted."
> (*San Diego Tribune* 10/5/ 89)— taken from Van Hoek (1997:124)

> *His* beloved old white convertible was in terrible shape, but *my stepfather* refused to get rid of it.   (*Reader's Digest* 2/89)
> — taken from Van Hoek (1997:124)

Two instances of this pattern (b) found in the AT:

[A004 89] " A lot of players dread thinking about *it* , but *that day* always comes in baseball when you have to go out and get a real job, " he said .

[A006 35] I can not promise *it* , but we will try for *$1,000*".

In her discussion of this pattern (b), Mittwoch (1983:133) states:

> ... the second conjunct represents the speaker's main point,.... It is therefore the appropriate point for the introduction of the new information represented by the antecedent.
> (Mittwoch 1983:133)

It can be said, at least, that pattern (b) ('S-pronominalization' and 'unequal conjoined structure') is likely to place focus on the elements of the second clause including the antecedent.

(c) Parenthetical construction

Van Hoek (1997:123) presents the following cataphora cases involving the cataphoric pronouns within the parenthetical constructions.

> It was 6:10 p.m. — almost time for *her* break — when ***Claudia Hawkins*** pulled her airport shuttle into Lot C.
> (*Los Angeles Times* 2/2/91)

> Today in Monte Carlo — the same princely locale where eight years ago *he* retired — *tennis legend Bjorn Borg* attempted a come back of his own.
> (*San Diego Tribune* 4/23/91)

> At the end of that time — and none knew who had started *it* — *a rumour* went through the jungle that there was better food and water to be found in such and such a valley.
> (Rudyard Kipling, *The Jungle Books*)

> This, the last of my books to have the benefit of *his* help and advice during its making, is dedicated to the memory of my husband, ***GEOFFREY LOFTS*** 1905-1948.
> (Norah Lofts, *Women in the Old Testament*, dedication page)

(d) The antecedent is included in the object of main verb

In the following examples, the cataphoric pronoun appears in a rather 'heavy' main subject NP, followed by the antecedent included in the object of main verb.

> An unsolved murder and *his* twin brother put ***Father D.*** in double jeopardy.
> (*TV Guide* 2/7/90, text of advertisement)

— taken from Van Hoek (1997:121)

An old flame's appearance at *his* high school reunion prompts *Jack (Morgan Stevens)* to dwell on precious moments of the past.
      (*TV Guide* 3/12/88) — taken from Van Hoek (1997:122)

Their dome-shaped heads and smaller ears set *Asiatic* apart from African elephants.
      (first line on the sign for the Asiatic elephant pen at the San Diego Zoo)
      — taken from Van Hoek (1997:122)

[A049 49] A slump in fourth quarter camera sales and continuing marketing problems with *its* instant movie system has dropped *Polaroid Corp.*'s 1979 earnings to less than half last year's levels.

(e) Right dislocation (Postponed identification)

Below is an instance of what is called 'right dislocation' or 'Postponed identification' that "involves placing a pro-form earlier in the sentence while the noun phrase to which it refers is placed finally as an amplificatory tag."(CGEL:1310).

[A047 67] *They* are shrouded in mystery , *these hockey playing visitors from the Soviet Union*, and there is irony in that .

(f) The reporter's supplying information in DS

In [A037 70] below, the reporter supplies additional information enclosed with the round brackets, (*Aikens' knee*), for readers to resolve the 3PP *it* in the DS context.

[A037 70] At the end of the day, *it (Aikens' knee)* looked a little puffy .

4-3-2. Cross-deictic coreference

Apart from the straightforward grammatical rule that a singular 1PP in a direct speech is usually coreferential with the subject of the reporting clause, there is not a formal rule to identify the coreferential relations in the cross-deictic reference.

(g) Plural 1PP in the initial DS has its antecedent in the reporting clause.

A plural 1PP in the initial DS can be recognised as cataphoric when a coreferential full NP appears in the reporting clause.[27]

[A029 43] "I thought *we* had the game won in regulation," said *North Carolina* Coach Dean Smith.

[A046 6] " None of *our* (previous) cases requires *us* to invalidate these reimbursements simply because they involve payments in cash, " Justice Byron R. White wrote for *the court* Wednesday.

(h) 3PP in the initial DS has its antecedent in the reporting clause.

A 3PP in the initial DS can be recognised as cataphoric when a coreferential full NP appears in the reporting clause.

[A051 80] "I saw that it was a traumatic experience for *her* to see that stone," said Mrs. Vrhovac, who was with *Mrs. Kovatch* when she first saw the inscription.

[A044 71] "I'm really keen to play *her* again," Barker said of her meeting with *the high school freshman from Lincolnshire , Ill.*

(i) Cross deictic cataphora involving a tag question in DS

[A057 39] "Not the easiest place to sign a treaty on, is *it*?" he said, running his hand over *the rough surface of the table.*

---

[27] Further examples are included in Appendix 4-3-2.

4-4. Summary and Conclusion

The corpus data of cataphora presented by Carden(1982), Kanzaki (1997), and Van Hoek (1997), as well as the AT data, include First-mention cases, which shows that personal pronouns can be used in reference not only to given/recoverable information but also to new information in written discourse. The AT data also indicate that a cataphoric reading is possible not only for the genuine cataphora cases (First-mention and Overriding types) but also for other Previously-mentioned cases.

The corpus data looked at in this section exhibit regular structural patterns of cataphora: the predominant initial-ADV pattern and 1PSP-initial-DS pattern, as well as some other patterns such as 'unequal conjoined structure' (Van Hoek, 1997:124). These frequent patterns may provide a conventional schema for the cataphoric reading of person pronouns. It may be the case that the more readers are exposed to these patterns, the more they become accustomed to the cataphoric interpretation of them.

Chapter 5: Generative Approaches to Cataphora

5-1. C-command coreferential constraint by Reinhart (1976, 1981, 1984)

5-1-0. Introduction

In generative grammar, several attempts have been made to capture the syntactic conditions which operate on coreferential relations between NPs within a sentence (Langacker 1969, Lakoff 1968, Reinhart 1976, 1981, 1984, Chomsky 1981). They are concerned with specifying the syntactic conditions in which an NP can or cannot be coreferential with the other NP within a sentence. One oft-cited, seemingly successful of such models is the syntactic 'domain' model defined by 'C-command',

initially proposed by Reinhart (1976, 1981, 1984). The notion of 'C-command' with some modification has been adopted as a key concept in GB theory (Chomsky 1981, 1985, 1986) and its variants in the generative framework. It has also been accepted in other fields of research, for example in cognitive approaches such as that of Ariel (1990).

In this section, I will discuss in detail coreferential constraints based on the syntactic 'domain' of 'C-command' (henceforth the 'C-command model') proposed by Reinhart (1976, 1981, 1984). Her analysis includes one of the main syntactic descriptions of cataphora (backward anaphora) involving initial (preposed) adverbial constituents: this is a typical cataphora pattern frequently occurring in the AT as well as in other corpus data (Carden, 1982; Van Hoek, 1996; Kanzaki, 1994).

Presumably partly because of the empirical difficulties of her original C-command model (presented in Reinhart,1976, 1981a, chapter 2-4 in 1984) and partly because of the need to take account of GB theory (Chomsky, 1981), Reinhart, in later publications, abandoned her model of sentence-level constraints on coreferential relations between NPs. Instead, she presented her model of sentence-level constraints on, what she called, 'bound anaphora'.

In this section, however, I shall discuss her original analysis for the following reasons:

(1) It is her original C-command model that has gained influence not only within the generative framework but also within the cognitive approaches such as the analysis of cataphora by Ariel (1990), which will be discussed in 7-4.

(2) Although Reinhart (1984:preface) says that her previous analysis "was incorrectly stated", she still claims, "in its empirical results, the analysis I propose now is largely

equivalent to my previous analysis" (ibid.). It is these "empirical results" that I am

mainly concerned with. For those who do not share theoretical backgrounds with each

other, such as generative linguists and myself, it is empirical data that usually provide

the common ground for their discussion.

(3) Even after abandoning her previous analysis, Reinhart (in chapter 7 of

Reinhart, 1984) still maintains the following points to which my arguments are

addressed.

(3-1) A distinction can be made between what she calls '**verb-phrasal**' PPs and

'sentential' PPs.

(3-2) There is a difference in syntactic position between preposed '**verb-phrasal**' PPs

and preposed 'sentential' PPs.

In the following illustration of Reinhart's C-command model, I will present

her work not in the same order as she presents her work in Reinhart (1983) but in the

order that makes clear how her arguments and claims are related to empirical data.

In 5-1-1, I will describe the concepts and terminologies of Reinhart's C-command

model, and clarify its empirical content and predictions. 5-1-2 presents initial

empirical problems that this model encounters in dealing with a typical pattern of

cataphora involving initial adverbials. 5-1-3 and 5-1-4 discuss Reinhart's solution to

the empirical problems presented in 5-1-2. 5-1-3 look at the distinction between

'**verb-phrasal**' PPs and 'sentential' PPs made by Reinhart. 5-1-4 examines her

'syntactic evidence' for differentiating the syntactic position of preposed '**verb-

phrasal**' PPs from that of preposed 'sentential' PPs. In 5-1-5, I will mention

Reinhart's misleading statements on the validity of her model and its consequences.

5-1-1. An overview of the coreferential constraints by C-command domain by Reinhart (1976, 1981, 1984)

At first, I will explain the terminology and concept of the coreferential constraints based on the syntactic 'domain' of 'C-command'(C-command model) proposed by Reinhart (1976, 1981, 1984).

Reinhart (1984:41-42) asserts that not only coreferential constraints, but all sentence-level interpretative rules conform to the general conditions stated in (I) and (II) below.

The general rule (I)

Sentence-level rules may operate on two given nodes only in the case that one of these nodes is in the domain of the other.

The general rule (II)

If a rule assigns node A some kind of prominence over node B, A must be a D(omain)-head of the domain containing B.

To illustrate these rules with Figure (5.1.1), the general rule (I) states that a sentence-level interpretative rule can operate between NP1 (positioned at node A) and NP2 (positioned at node B) only if either NP1 (node A) is in the 'domain' of NP2 (node B)

(5.1.1)

```
                    S
                  /   \
                /       \
       NP1 (node A)      VP
                        /  \
                      /      \
                     V       NP2
                           (node B)
```
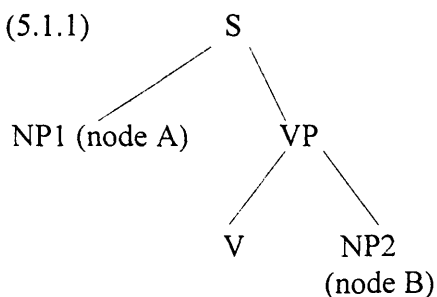
or NP2 is in the 'domain' of NP1, but it cannot operate if neither is in the 'domain' of the other. In the latter cases, their coreferential interpretation "depends on pragmatic, rather than sentence-level interpretative considerations." (1984:42). Hence the definition of 'domain' determines the extent to which a sentence-level interpretative rule can be applied. The notion of 'C-command' is proposed as one of such definitions of syntactic 'domain' within which sentence-level interpretative rules can operate. (36) gives the definitions of C-command (Reinhart,1984:41) with its illustration in Figure (5.1.2).

(36) Node A C-commands node B if the branching node $\alpha 1$ (most immediately dominating A) either dominates B (above figure), or is immediately dominated by node $\alpha 2$ which dominates B, and $\alpha 2$ is of the same category type as $\alpha 1$ (figure below).

(5.1.2)
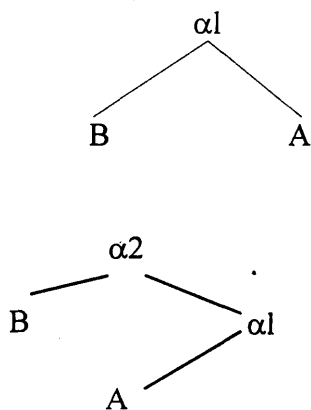
(37) below defines the domain of a node based on C-command:

(37) The domain of a node A consists of all and only the nodes C-commanded by A.

Figure (5.1.3) below serves to illustrate both the definitions C-command (36) and the 'domain of a node' (37). In figure (5.1.3), NP1 (subject) C-commands NP2 (object) since the node S, most immediately dominating NP1 (subject), dominates VP that dominates NP2 (object). Hence NP2 (object) is said to be 'in the domain of NP1 (subject)'.

(5.1.3)
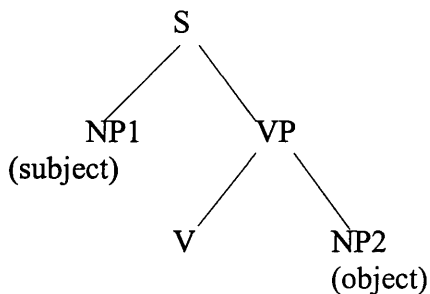
The domain of node NP1 (subject) consists of all nodes of the tree: VP, V, NP2, NP1 itself and S (since a node C-commands its immediate mother)[1].

The domain of node S and VP also consists of all nodes of the tree.

NP2 (object), on the other hand, does not C-command NP1 (subject) since VP, most immediately dominating NP2 (object), does not dominate NP1 (subject), and VP is a different category type from S that dominates NP1 (subject).

Hence NP1 (subject) is not in the domain of NP2 (object). The domain of NP2 (object) consists of V, NP2 itself and its immediate 'mother' VP. The domain of V is exactly the same as that of NP2.

---

[1] Reinhart (1981:614, 1984:21) states that a given node C-commands itself and also C-commands its immediate 'mother' node that immediately dominates the given node. That means, in (7.2.3), both NP1 and VP C-commands S as well as V and NP2 C-commands VP. Although this definition has several problems, since they are not directly relevant to the current discussion of cataphora, I will not discuss them here.

Figure (5.1.4) serves to illustrate how Reinhart's C-command model can be applied to the typical cataphora construction involving initial adverbial constituents (PP)[29].

In (5.1.4), NP1 (the subject of the main clause, henceforth 'main subject') C-commands PP (initial adverbial constituent) at COMP position[30]. This is so due to the definition in (36): S1 (dominating main subject NP1) is immediately dominated by S2, which dominates PP including NP2, and S2 is of the same category type as S1.

(5.1.4)

```
                    S2
                  /    \
                /        \
              /            \
          PP(COMP)          S1
          /     \          /    \
        /         \       /        \
      P           NP2    NP1         VP
                      (Main subject)
```
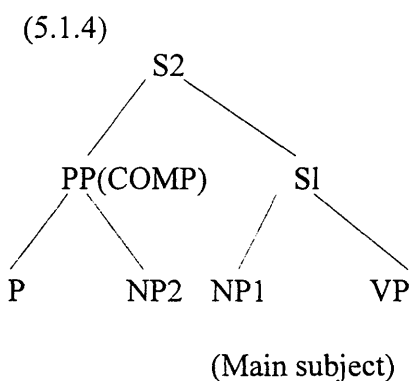
Hence NP2 (under PP) is in the domain of main subject NP1. NP2, on the other hand, does not C-command NP1 (main subject) since PP, most immediately dominating NP2, does not dominate NP1 (main subject), and PP is a different category type from S2 that dominates NP1. Hence NP1 (main subject) is not in the domain of NP2.

---

[29] Reinhart uses the symbol 'PP' not only for prepositional phrases but also initial adverbial clauses. For example, she uses "P and S" to refer to "*after-*, *when-*, *because-*clauses" (1984:75). She also uses the symbol 'PP' for initial non-finite clauses as in the following example (1984:153).

Thinking about his problems Siegfried got depressed and Felix (did) too.

So I will assume all the PPs in her articles refer to initial adverbial constituents in general.

[30] According to Reinhart (1984:68), COMP is the syntactic position for 'complementizer' such as wh-words, Questions, Imperative markers, and preposed PPs.

Based on the syntactic domain defined by C-command, the general rule (II) will determine the coreferential constraints between NPs within the domain.

<u>The general rule (II)</u>

If a rule assigns node A some kind of prominence over node B, A must be a D(omain)-head of the domain containing B.

The terms 'D(omain)-head' and 'prominence over' need to be explained here. A 'D(omain)-head' of domain α is a node which "C-commands all and only the nodes in α" (1984:23). To illustrate this with (5.1.3), asexplained before, the domain of NP1 (subject) is the entire set of nodes of the tree. Hence NP1 (subject) is said to be a 'D(omain)-head' of the domain which consists of the whole set of nodes.

(5.1.3) -- repeated --

```
            S
          /    `.
        /         `.
      /              `.
    NP1               VP
 (subject)          /    \
                  /         \
                /             \
              V                NP2
                             (object)
```
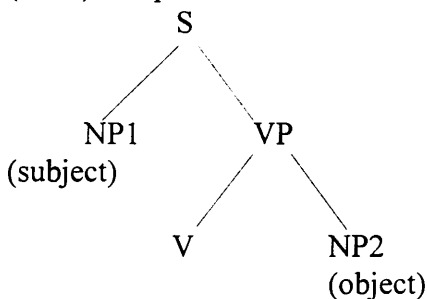
Since VP and S also share the same domain as NP1, all three nodes are D-heads of the same domain consisting of all nodes in the tree.

The term 'prominence' in the rule (II) is meant to be associated with "referential independence" (1984:42). A full NP (non-pronoun), according to Reinhart, is 'referentially independent' while a pronoun is not. Pronouns are 'referentially dependent' in that for their interpretation, pronouns need 'referentially independent' full NPs (antecedents). Hence when NP1 and NP2 are coreferential and

119

NP1 C-commands NP2, the general rule (II) states NP2 must be a pronoun.[31] The

coreferential interpretation rule is stated in (40).

(40) A given NP must be interpreted as non-coreferential with any distinct

non-pronoun in its C-command domain.

(1984:43)

To illustrate this coreferential constraint using Reinhart's cataphora examples

(45)(46)(47) with figure (5.1.5):

(5.1.5)

```
                            S'
                  _____
           PP (COMP)              S
            /    \          _____
           P     NP2      NP1           VP
           |      |    (main subject)   / \
           |      |         |          /   \
(45) Near  him         Dan              saw a snake.
(46)*Near  Dan         he               saw a snake.
(47)*Near  Dan         Dan              saw a snake.
```

(Reinhart, 1984:44)
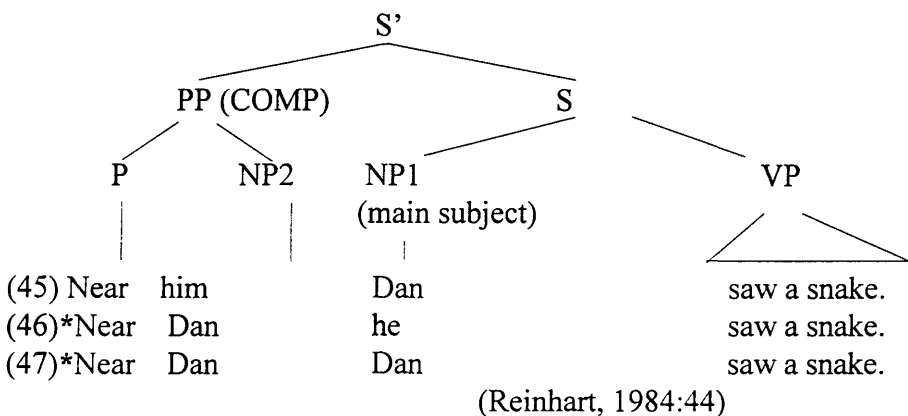
[Note] The asterisk (*) indicates Reinhart's judgement on ungrammaticality of
the sentence.

In (5.1.5), the main subject (NP1) has NP2 in its domain (as explained before with

(5.1.4)). If the main subject NP1 is coreferential with NP2, the rule (40) states that

NP2 has to be a pronoun. The sentence (45) satisfies this condition whereas (46) and

(47) do not (hence marked with '*').

In order to grasp the empirical prediction of the C-command model and clarify

the supportive evidence as well as counter-examples to the model, Table 5.1.6 below

---

[31] Reinhart does not clearly state that NP1 must be a full NP (antecedent) or it can be
a pronoun.

enumerates all the possible patterns of relation between two NPs in terms of C-command domain. For each pattern, the coreference prediction made by the C-command model and counter example cases are stated.

[Notations of Table 5.1.6]

* 'Pron' stands for 'pronoun', and 'fNP' stands for 'full NP'.

* 'NP1 → NP2' indicates that NP1 C-commands NP2.

* 'NP1 ↔ NP2' indicates that both NP1 and NP2 C-command each other.

* 'NP1 ? NP2' indicates that neither NP1 nor NP2 C-command each other.

* 'NP1 = NP2' indicates that NP1 is coreferential with NP2.

* 'NP1 ≠ NP2' indicates that NP1 is not coreferential with NP2.

* 'out of scope' means that C-command model cannot be applied to this case. It is

   outside the scope of application of the C-command model.

Table 5.1.6: Possible patterns of relation between two NPs in terms of C-command domain

Relation pattern	C-command relation between two NPs	coreference prediction by C-command model	counter example cases
Pattern 1	Pron → Pron	no prediction	?
Pattern 2	Pron ↔ Pron	Pron ≠ Pron	Pron = Pron
Pattern 3	Pron ? Pron	out of scope	
Pattern 4	fNP → fNP	fNP ≠ fNP	fNP = fNP
Pattern 5	fNP ↔ fNP	fNP ≠ fNP	fNP = fNP
Pattern 6	fNP ? fNP	out of scope	
Pattern 7	Pron → fNP	Pron ≠ fNP	Pron = fNP
Pattern 8	Pron ← fNP	Pron = fNP Pron ≠ fNP	no
Pattern 9	Pron ↔ fNP	Pron ≠ fNP	Pron = fNP
Pattern 10	Pron ? fNP	out of scope	

In the typical cataphora construction involving an initial adverbial (5.1.4), the NP2

within the initial adverbial (PP) is C-commanded by the main subject NP,

(5.1.4) -- repeated

```
            S2
          /    \
        /        \
   PP(COMP)       S1
    /  \         /  \
   /    \       /    \
  P     NP2   NP1     VP
              (Main subject)
```
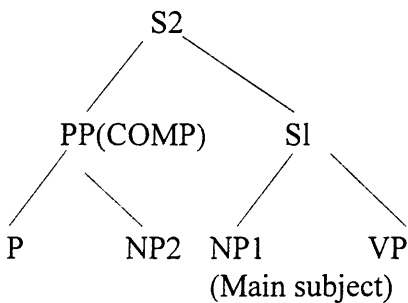
but the main subject (NP1) is not C-commanded by the NP within the initial PP

(NP2). In this respect, the C-command relation between the main subject and the NP

within initial PP is one-directional (Main subject → NP in PP). Hence only Patterns 4,

7, and 8 in Table 5.1.6 (listed below in (5.1.7)) are relevant to the current discussion

of a typical cataphora involving initial adverbials.

Table (5.1.7): Patterns of relation between two NPs in terms of C-command domain
relevant to initial-ADV cataphora

Relation pattern	C-command relation between two NPs	coreference prediction by C-command model	counter example cases
Pattern 4	fNP → fNP	fNP ≠ fNP	fNP = fNP
Pattern 7	Pron → fNP	Pron ≠ fNP	Pron = fNP
Pattern 8	Pron ← fNP	Pron = fNP Pron ≠ fNP	none

From Table (5.1.7), the empirical contents of the C-command model can be stated as

below:

* Supporting evidence for Reinhart's C-command model is given by an instance

where:

(5.1.7.a) A full NP C-commands its coreferential pronoun, but the latter does not

C-command the former. (Pattern 8: full NP → pronoun)

* Counter-examples for the C-command model are given by the occurrences of

coreferential two NPs :

(5.1.7.b) A full NP C-commands its coreferential a full NP, but the latter does not

C-command the former. (Pattern 4: full NP → full NP)

(5.1.7.c) A pronoun C-commands its coreferential a full NP, but the latter does not

C-command the former. (Pattern 7: pronoun → full NP)

Based on the empirical contents of Reinhart's C-command model described above, in

the next section, I will look at empirical problems of the model in dealing with

cataphora data in AT.

5-1-2. Initial empirical problems for Reinhart's C-command model

One of the typical patterns of cataphora found in AT is that where a pronoun in the initial adverbial is followed by its coreferential main subject NP (provided that the pronoun does not have a preceding coreferential NP within the initial adverbial). In AT, this type of cataphora construction appears 46 times among 133 occurrences of all cataphora examples (34 percent) (see 4-2-2). If we assume that the initial adverbial constituents are positioned at COMP (5.1.4) hence C-commanded bythe main subject NPs, those 54 examples appear to provide supporting evidence for Reinhart's C-command model, as stated in (5.1.7.a).

(5.1.4) --- repeated

```
                S2
              /    \
             /      \
      PP(COMP)       Sl
       /  \         /  \
      /    \       /    \
     P    NP2    NP1     VP
             (Main subject)
```
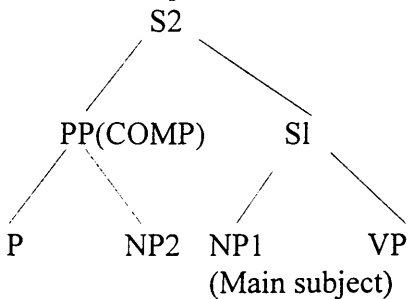
This assumption, however, also results in providing counter-examples to her model. As mentioned in the previous section, there are two types of counter-example cases for Reinhart's C-command model:

(5.1.7.b) A full NP C-commands a coreferential full NP, but the latter does not C-command the former. (Pattern 4: full NP → full NP)

(5.1.7.c) A pronoun C-commands a coreferential full NP, but the latter does not C-command the former. (Pattern 7: pronoun → full NP)

If we assume that initial adverbials are C-commanded by main subjects, these two

types of counter-examples often can be found in AT as well as in other corpora.[32]

One type of counter-example (5.1.7.b), <full NP → full NP> pattern appears in [A013

42] and [A016 43], where the main subject full NP C-commands its coreferential full

NP.

> [Note] The main subjects and its coreferentail pronoun in the initial PP is
> typed in bold-italics.

[A013 42] Though *the plan* does not spell out implications of the revised
mathematics, *the new system* apparently would help some of the state's
smaller hospitals defend their continued existence.

[A016 43] After a June 1964 plane crash in Southampton in which ***Democratic U.S.
Sen. Edward Kennedy*** broke his back , ***Kennedy*** credited Corriden with
saving his life .

(5.1.8) Tree diagram for [A013 42] assuming that initial (preposed) PPs are positioned
at COMP.
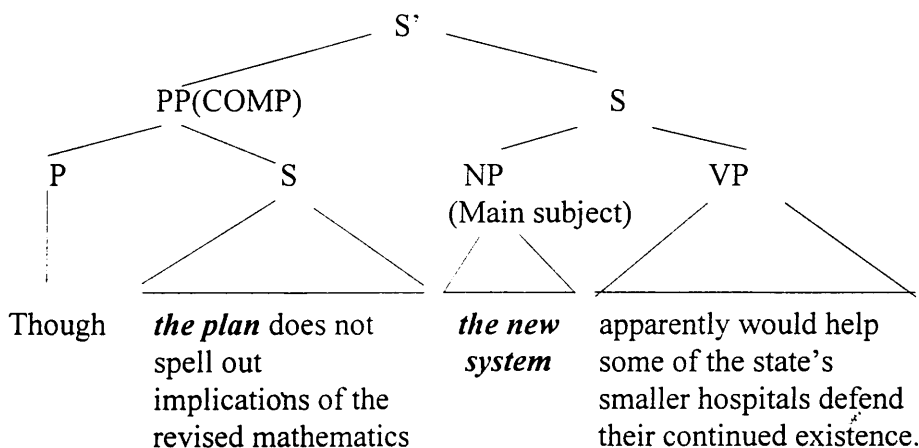
Another type of counter-example, (5.1.7.c) <pronoun → full NP>, can be found

in [A015 54] and [A012 46] below, where the main subject pronoun C-commands its

coreferential full NP in the initial adverbial clause.

---

[32] In fact there are many instances of the counter-examples, but only a couple of
examples are listed for each type.

[A015 54]
 When *Cleveland firemen* arrived at the house, *they* found Robert lying behind the bed in Mrs. Dumas ' upstairs bedroom.

[A012 46]
 When *Mike Calise* heads south next month, *he*'ll have a special item packed in with his baseball equipment - a boxer's glove.

[A025 46]
 If *Huck Finn* lived in Connecticut today, *he* would probably be a student in an alternative school.

(5.1.9) Tree diagram for [A015 54] assuming that initial (preposed) PP are positioned at COMP.
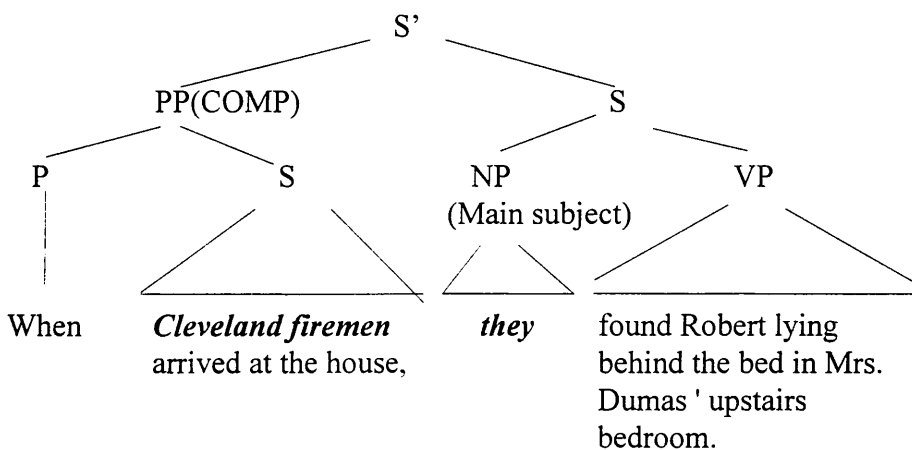
Van Hoek (1997:113) also collected 100 example of <pronoun → full NP> type

(termed 'repeated identification') in which a pronoun appears at the subject of main

clause while the coreferential lexical NP (antecedent) appears within the initial

adverbial constituent, as below.

a. As *T. J.* returned to his seat, *he* moved his chair close to the last row of students. (Reader's Digest 4/91)

b. When *a Cockney lad named Maurice Joseph Micklewhite, Jr.* was 19, *he* was sent to the Korean war. (Reader's Digest 11 / 91)

This type is also mentioned in a descriptive grammar:

> Cataphoric reference occurs less frequently, and under limited conditions.
> Where it does occur, anaphoric reference is also possible ..
>
> [1] Before *he* joined the Navy, *Gerald* made peace with his family.
> [1a] Before *Gerald* joined the Navy, *he* made peace with his family.
>   (Quirk et al, CGEL:351)

In the generative framework, also the same observation has been made, as Kuno and

Takami (1993:157) noted,

> It has also been known since Lakoff (1968) that sentences ... involving
> fronting of complex NPs or clausal constituents allow both forward
> and backward pronominalization ..

To cope with this initial empirical problem with her C-command model,

Reinhart firstly introduces a distinction between 'sentential' PPs and 'verb-phrasal'

PPs; secondly she claims that only preposed **verb-phrasal** PPs are C-commanded by

the main subject, whereas preposed **sentential PP**s are not C-commanded by the main

subject; and thirdly she identifies the initial PPs in the seeming counter-examples to

her C-command model as initial **sentential** PPs. Hence they are not counter-examples

but out of the scope of the application of her model. Figure (5.1.10) contrasts a **verb-**

phrasal PP at initial position with a **sentential PP** at initial position, according to

Reinhart's analysis.

(5.1.10)   ---   (42) in Reinhart(1984:70)

verb-phrasal PP at initial position   **sentential PP** at initial position
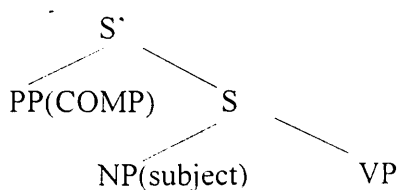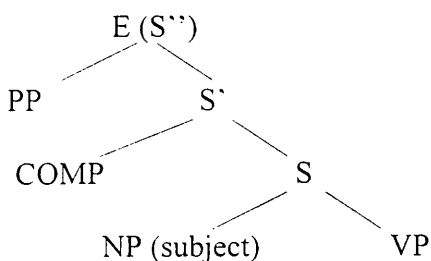
According to Reinhart(1984:70), the main subject NP C-commands the initial PP at COMP (Left tree), but does not C-command the initial PP at the position immediately dominated by E (S'') (Right tree in 5.1.10). Hence the C-command model does not operate on the **sentential PP** at initial position (Right tree in 5.1.10). Reinhart (1984:75) says this "explains why coreference can go both ways when they (**sentential PPs**) are in initial position". In other words, when a **sentential PP** at initial position includes a NP which is coreferential with the main subject, the NP included in PP can be either a pronoun or a full NP. Thus, in Reinhart's examples below (5.1.11), not only (a) but also (b) (c) are possible.

(5.1.11)

```
                    E (S")
                   /      \
              PP              S'
            /    \          /    \
         P    NP    COMP         S
         |    /\                /   \
         |   /  \             NP      VP
         |  /    \            |      /    \
```

(a) In *his* office,	*Ben*	is an absolute dictator.
(b) In *Ben*'s office	*he*	is an absolute dictator.
(c) In *Ben*'s office	*Ben*	is an absolute dictator.

(example 47a,b,c in 1984:73)
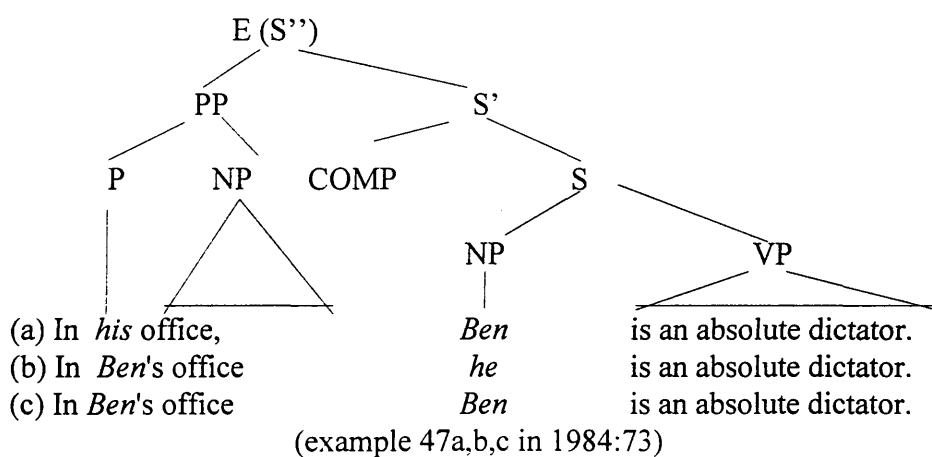
This is because the initial (preposed) PPs in these sentences are no longer positioned at COMP which is C-commanded by main subject, but positioned at the higher node (dominated by E or S'') which is not C-commanded by the main subject.

Following Reinhart's argument, those corpus examples presented as counter-example cases (<full NP → full NP> and <pronoun → full NP> type) are considered to involve not preposed **verb-phrasal** PPs but preposed **sentential PPs**; and the

pronouns within PPs are not C-commanded by their coreferential main subjects; hence neither of main subject NPs nor NPs within initial PP are C-commanding each other; in other words, those seeming counter-examples from corpora are not counter-examples but out of the scope of applying her model. Her model does not operate on those examples. By differentiating syntactic position between initial **sentential PP** and initial **verb-phrasal** PP, Reinhart rescues the C-command model from the counter-examples of AT presented before. It should be noted, however, that her model now can operate only on those cataphora (backward anaphora) cases involving initial **verb-phrasal** PP. It can no longer be applied to those cases involving initial **sentential PPs**.

Before looking at Reinhart's argument on the distinction between **verb-phrasal** and **sentential PPs**, I would like to focus on the inadequate statements in term of the scope of the application of Reinhart's C-command model made by herself as well as others.

5-1-3. Inadequate statements on the validity of the C-command model.

As noted before with Table (5.1.6), Reinhart's C-command coreferential model cannot operate between the two NPs when neither of them C-commands the other, regardless of whether either or both of them are pronouns or full NPs. 'Out of scope' cases in Table (5.1.6) --- repeated

Relation pattern	C-command relation between two NPs	coreference prediction by C-command model	counter example cases
Pattern 3	Pron ? Pron	out of scope	
Pattern 6	fNP ? fNP	out of scope	
Pattern 10	Pron ? fNP	out of scope	

Reinhart says (1984:42):

> there are no sentence-level restrictions on the coreference interpretation of two NPs when neither is in the domain of the other. In such cases, coreference should always be permitted syntactically, regardless of whether one of the NPs in question is a pronoun. Whether the two NPs are assigned the same or different references in such cases depends on pragmatic, rather than sentence-level consideration.

Following this, it is clear that Reinhart's C-command coreferential model does not operate on the coreferential relation between an NP in initial **sentential PP** and the main subject NP, since neither of those NPs C-commands the other. This case is beyond the scope of the application of her C-command coreferential model, and it provides neither supporting evidence nor counter evidence to the model. According to Reinhart, the coreference "in such cases depends on pragmatic, rather than sentence-level considerations." (ibid.). It is, therefore, inadequate to say that her C-command

model 'correctly allows' or 'correctly permits' the coreference involving initial

sentential PPs since the model cannot make any coreference-prediction to be judged

correct or not for such cases.

Such inadequate statements, however, can be found in Reinhart's account as

in:

...(72b) Near him, Don's mother found a gun.

The C-command restriction ... also, correctly, permits coreference in (72b), since the pronoun is dominated by a (locative) PP, and, therefore, it does not C-command the antecedent.

(Reinhart, 1981:631, 1984:54)

The tree diagram for (72b) would be:

(5.1.12) Tree for (72b) in Reinhart, 1984:54

```
                      S'
            _____/  _____
          PP                    S
        /   \              ___/    _____
       P     NP        NP(subject)         VP
       |     |          /      \          /    \
       |     |    Det(SPEC)     N'       /      \
       |     |        |         |       /        \
     Near   him,    Don's     mother  found a gun
```
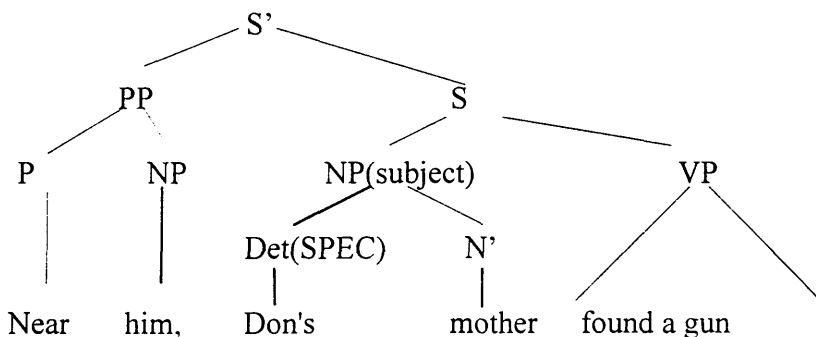
It is clear that in (72b) neither *him* in the initial PP nor *Don* (under SPEC[33]) C-

commands the other. This example is out of the scope of the application of her C-

command model, and the model cannot make any prediction of coreference for these

NPs. Hence for this example, it is inadequate to say that the C-command model

'correctly permits coreference' between *him* in the initial PP nor *Don* (under SPEC),

simply because the model cannot make any coreference-prediction to be judged

---

[33] SPEC is the syntactic position for 'specifier' (e.g. determiners) in GB theory. (Haegeman,1991:119)

correct or not for these NPs. If it is allowed to say that the model 'correctly permits coreference' to those NPs, then it must also be allowed to say that the model 'correctly permits non-coreference' between those NPs since both coreference and non-coreference readings are possible for (72b) depending on the context.

In fact, it is this kind of statement that is not only inadequate and misleading but also causes theoretical confusion. It ironically provides good reasons for the criticisms against Reinhart's account itself. For instance, Kuno (1987) argues that Reinhart's C-command analysis "**incorrectly predicts**" the acceptability of coreference for those NPs having **no C-command relation** with each other:

> Reinhart's C-command analysis, however, ... erroneously predicts that (6.8b) should be acceptable:
>
> (6.8) b. *Near John*(NP1), the investigator believed that he*i* (NP2) found a snake.
>
> ... In the above structure, NP2 no longer C-commands NP1,. Therefore, the C-command analysis incorrectly predicts that (6.8b) should be acceptable.

It should be stressed, however, that Reinhart's C-command coreferential model cannot operate on coreference relations outside the domain it defines. The example cited above cannot be a counter-example to Reinhart's C-command model because such examples are beyond the scope of the application of the model. Her model does not make any prediction to account for this example.

132

A similar point can be made about Takami(1985)'s criticisms of Reinhart's C-command model below:

(50) a. * You should ask *him* for some money, for *John* is a millionaire. 4
     b. * You can recognize *her* easily, because *Mary* wears big (V) glasses.
(51) a. * John likes *her*, while *Mary* hasn't even thought of him.
     b. * You can ask *him* anything since *John* is a real book- (V3) worm
(52) a. * Mary kissed *him* when *John* came to her house.

> ... the pronoun NP1 does not C-command its antecedent in each PP of (50)-(52). Hence both the C-command constraint and the binding theory incorrectly permit coreference in (50)-(52).
>
> (Takami, 1985: 286)

Takami criticises Reinhart's model saying that her model "**incorrectly permit coreference**" for those cases where there is **no C-commanding relation** between NPs. This criticism is not adequate for the same reason as Kuno's criticism.

The same point also can be made to Carden's criticism to Reinhart:

(20)   a. Mary showed *John*'s picture to *him*.
             (John = him)
         b. Mary spoke to *him* in **Ben**'s office.
             (him ≠ Ben)

> If dative PPs count as bounding nodes, the Surface Model (* Reinhart's model) will correctly permit the coreference in (20a), but incorrectly also permit coreference in (20b).
>
> (Carden, 1986:328):

If dative PPs count as bounding nodes[34], the 3PP *him* in (20a) and (20b) can C-command only the preposition *to*. Hence in (20a), neither the NP *John* (which is not in an argument position but an oblique position) nor 3PP *him* C-commands the other. The same is true between *him* and *Ben* in (20b). Hence both (20a) and (20b) are simply out of the scope of Reinhart's C-command model, and the model does not

---

[34] For the definition of 'bounding node', see 5-2-2.

predict or 'permit' any coreference or non-coreference between *John* and *him* in (20a) and between *him* and *Ben* in (20b). It is, therefore, inadequate to say, as Carden does, that Reinhart's model will "correctly permit the coreference in (20a)" or "incorrectly also permit coreference in (20b)".

It is now clear that a misunderstanding of the scope of the application of Reinhart's C-command model can be observed in the above criticisms made by Kuno (1987), Takami (1985), and Carden (1986).

Their misunderstanding seems to be triggered by an inadequate statement made by Reinhart herself (1984:54), which shows a misunderstanding of the scope of the application of her own model.

In the next section, I will look at Reinhart's arguments on the distinction in syntactic position between initial **verb-phrasal** and initial **sentential PPs**.

5-1-4. Reinhart's distinction in the syntactic position between initial sentential PP and initial verb-phrasal PP.

As mentioned in 5-1-2, in order to cope with the initial empirical problems of the C-command model, what Reinhart does is the following:

(1) firstly she introduces a distinction between 'sentential' PPs and 'verb-phrasal' PPs

(2) secondly she claims that only preposed verb-phrasal PPs are syntactically positioned at COMP and therefore C-commanded by the main subject, whereas preposed sentential PPs are positioned at higher nodes and are not C-commanded by the main subject.

This second step makes it possible to identify the counter-examples to her C-command model as preposed sentential PPs, so that they no longer breach the C-command coreferential constraints. This section focuses on the first step, the distinction between 'sentential' PPs and 'verb-phrasal' PPs, comparing it with the classification of adverbials in a descriptive grammar CGEL (Quirk et al 1985).

It has been widely recognised that some kinds of adverbials are closely integrated into the verb phrase, while some other adverbials modify the sentence as a whole being loosely integrated into the clause or verb phrase. The former type is often termed 'restrictive' and the latter type 'non-restrictive'.

A descriptive grammar such as CGEL (Quirk et al,1985:1074f) classifies adverbials as below:

(5.1.13)

```
                              Adverbials
        ┌─────────────┬──────────┴──────────┬─────────────┐
   Subjuncts        Adjuncts            Disjuncts      Conjuncts
              ┌──────────┴──────────┐
        Predication           Sentence Adjuncts
         Adjuncts
    ┌─────────┴─────────┐
Obligatory           Optional
Predication Adjuncts  Predication
                      Adjuncts
```
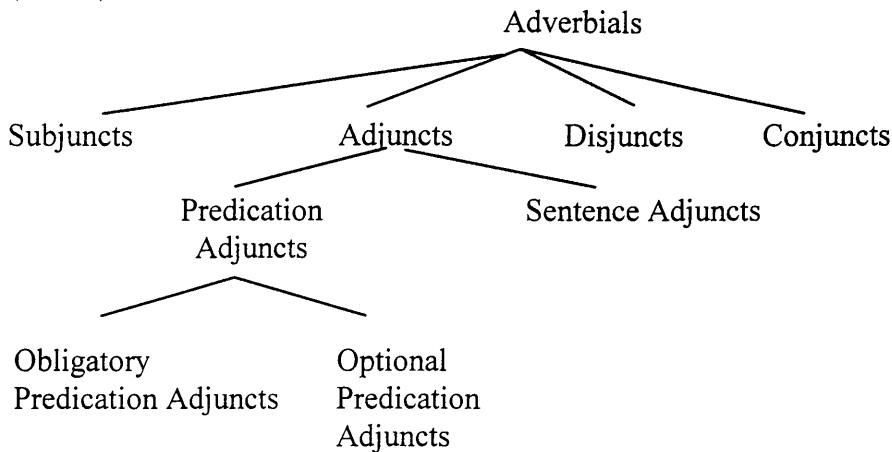
In CGEL's framework, the 'restrictive' VP-integrated type approximately corresponds

to **obligatory predication adjuncts** as well as part of the category of **optional**

predication adjuncts, and the 'non-restrictive' less-VP-integrated type corresponds

to **disjuncts, conjuncts, sentence adjuncts** and part of the category of **optional**

predication adjuncts. What Reinhart calls **verb-phrasal** PPs appear to correspond to

the former type, and **sentential PPs** to the latter type. There are, however, notable

differences between CGEL's and Reinhart's accounts. Firstly, Reinhart makes a clear-

cut distinction between **verb-phrasal** PPs and **sentential PPs** in a particular syntactic

position, viz when PPs are initially positioned. CGEL(1985:52), on the other hand,

assumes a spectrum rather than a clear-cut distinction between the two types of

adverbials, as stated below.

> Our position, however, is that adverbials represent a spectrum of types, the
> most central of which, because of their obligatoriness and relative immobility,
> resemble complements. ...
> At the other end of the spectrum, there are elements which are frequently
> called SENTENCE ADVERBIALS, because they tend to qualify, by their
> meaning, a whole sentence or clause, rather than just part of a clause (such as a
> verb, or a verb and object)
>
> (CGEL(Quirk et al,1985):52)

136

Secondly, according to Reinhart, whether a PP can occur in the initial position is determined by syntactic factors: (i) whether the PP is **verb-phrasal** or sentential and (ii) whether the sentence involves a wh-question or yes-no question or an imperative (explained in 5-1-5). According to CGEL, on the other hand, there are various conditions that influence the initial occurrence of adverbials: semantics of adverbials (time/location), end-focus, end-weight, clarity requirements (avoiding ambiguity), rhetorical requirements and so on. Thirdly, Reinhart's **verb-phrasal** PPs can take place in the initial position grammatically so long as the sentences does not involve a wh-question or yes-no question or an imperative. According to CGEL (p.1074), on the other hand, **predication adjuncts** are hardly ever positioned initially. An exception is for the rhetorical purpose of balancing two clauses as in:

> In London I was born and in London I shall die.
> (CGEL:1378)

Other exceptions may involve End-focus, End-weight, or clarity requirements (avoiding ambiguity).

CGEL (p.1627) also notes that since predication adjuncts "are strongly integrated within clause structure, they cannot be followed by a comma" when they are preposed. Hence according to CGEL(p.1627) the following sentence including a comma between the initial adverbial and the main clause is not acceptable.

> *To the very top of mountain, they climbed.
> taken from Note[a], CGEL:1627

In Reinhart's account, on the other hand, most of the examples of preposed **verb-phrasal** PPs are immediately followed by commas before the main clauses. Regarding the differences between Reinhart's account and CGEL's descriptions of

adverbials, it is hard to identify the status of what Reinhart calls **verb-phrasal** PPs

within the CGEL framework.

5-1-5 Reinhart's solution (II): Differentiating syntactic position between initial sentential PP and verb-phrasal PP

The second step of Reinhart's solution for the initial empirical problems is to differentiate the syntactic positions of initial **verb-phrasal** PP and initial **sentential PP**. This differentiation makes it possible to 'immunise' her C-command model against cataphoric counter-examples by identifying these counter-examples as initial **sentential PPs**; hence they can no longer be considered either counter-evidence or supportive evidence. Thus Reinhart's solution to the initial empirical problems is a theory-internal one, which is valid only within the generative framework. It appears to be adhoc for those who do not share the same framework as Reinhart, and it arises a question: how can you tell that initial **sentential PPs** are not positioned at the same node (COMP) as initial **verb-phrasal** PPs? On the surface, no structural difference appears to exist between initial **sentential PPs** and initial **verb-phrasal** PPs. Both types of PPs, at least, can be represented by the same structure in the skeleton parsing scheme([P P] [N N] [V V]) as shown with Reinhart's examples below.

> **Sentential** PP   (Example 47b in Reinhart,1984:73)
> [P In *his* office P] , [N *Ben* N] [V is an absolute dictator V] .
>
> **Verb-phrasal** PP   (Example 18a in Reinhart,1984:34)
> [P Near him P] , [N Dan N] [V saw a snake V].

What Reinhart needs to show is the empirical justification for the different treatment of initial **verb-phrasal** PPs and initial **sentential PPs**.

The justification presented by Reinhart (1982:68-69) is as follows: according to the COMP theory, there is only one COMP position in the sentence. This position can be occupied by only one of four constituents: a wh-word, a Q(uestion), an

Imp(erative) marker, or a preposed **verb-phrasal** PP, as shown in Figure (5.1.14).

(5.1.14)

```
                        E (S'')
              /                    \
     ┌──────────────┐               S'
     │ Sentential PP │          /        \
     └──────────────┘      COMP            S
                          /            /      \
  ┌──────────────┐       /     NP (main subject)   VP
  │ Q(uestion)   │──┐   /
  ├──────────────┤  │  /
  │ Imp(erative) │  │ /
  ├──────────────┤  │/
  │ wh-word      │ /
  ├──────────────┤/
  │ Verb-phrasal PP │
  └──────────────┘
```
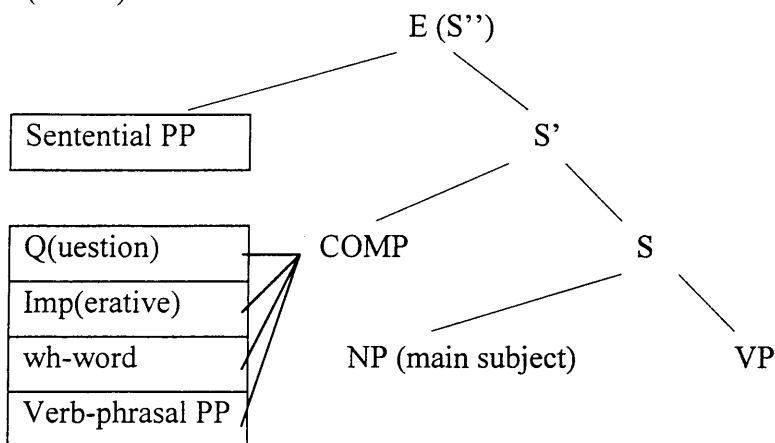
None of the four can co-occur with any of the others at the initial position. Hence when the COMP position is occupied either by a wh-word, or a Q(uestion) or an Imp(erative) marker, a **verb-phrasal** PP can no longer take place initially. **Sentential PPs**, on the other hand, do not occupy the COMP position and can take place initially together with a wh-word, or a Q(uestion) or Imp(erative) marker.

Reinhart(1984:68-69) presents the following 'empirical evidence' to support her claim. All the eight examples presented below involves what she calls initial **verb-phrasal** PP co-occurring with either wh-words, Q(uestion)s or Imp(erative) markers; hence, according to Reinhart, they are ungrammatical and marked with asterisks (*).

[Wh-words]

    (34b) * In Ben's picture of her, what did she find?
    (35b) * In Detroit, who did the gangsters kill?
    (36b) * In front of Felix, who holds a candle?
    (37b) * With her new boss, why does(n't) Rosa argue?

[Questions]
    (38b) * In Ben's picture, did Rosa find a scratch?
    (39b) * With her boss, does she argue?

[Imperatives]
>   (40b) * In my next picture, find a scratch, if you can!
>   (41b) * With your boss, stop arguing!

To examine her claim and this 'evidence', I will conduct an informant test with fifteen native speakers, investigating the acceptability of these eight examples. In preparing examples for this kind of acceptability test, it is important to make clear the objectives of the test and relevant check-points for the examples. For instance, if the objective is to examine the use of past simple forms in comparison with the present perfect tense, among the following examples:

>   (a) You haven't got the ticket yet?
>   (b) You didn't get ticket yet?

(a) is appropriate to be used for the test but (b) is not, since (b) contains an ungrammatical element (the singular noun 'ticket' does not have a determiner) that is not relevant to the objective of this test. Some informants may judge (b) as ungrammatical because of this irrelevant ungrammatical element ('noise' -- so to speak). Because such 'noise' damage the credibility of tests, they must be avoided as much as possible.

As to Reinhart's case, the claim she makes with reference to a wh-word, a Q(uestion), an Imp(erative), and an initial **verb-phrasal** PP, is that none of them can co-occur with any of the others at in initial position. According to this claim, Reinhart's eight examples (34b-41b) must be ungrammatical since all the preposed PPs are initial **verb-phrasal** and they occur either in a wh-question, or in a yes-no question, or in an imperative sentence. The factors relevant to Reinhart's grammaticality judgements are:

1. Position of PP (initial or other position)

2. Type of sentence (wh-question or yes-no-question or imperative or others)

3. Lexical combination of verb and preposition. There are four combinations of verb and preposition in the eight examples: *find* and locative[35] *in*, *hold* and locative *in front of*, *argue* and accompaniment *with*, and *kill* and locative *in*.

It must be emphasised that other factors (such as lexical, grammatical elements of NPs or schematic knowledge) should not be allowed to influence grammatical judgements on these sentences because such factors are irrelevant to her claim. In other words, the examples must not contain 'noise' if they are to provide 'empirical evidence' for her claim. If the deficiencies caused by irrelevant factors influence acceptability judgements on the sentences, it will damage the credibility of the test, and it will be difficult to present them as 'empirical evidence' for her claim. To make sure that the three relevant factors do not impair grammaticality in the alternatives I have included an alternative sentence to each Reinhart's example in the informant test. The objective of the informant test is to examine whether what Reinhart calls **verb-phrasal** PPs really cannot take place with wh-questions, or yes-no questions, or imperative sentences.[36]

.

---

[35] I am following Reinhart's manner of specifying types of PPs as in:

> locatives (e.g. in NP) are obligatory verb-phrasal following verbs like dwell, reside, or put. Similarly the verb flirt requires a with-phrase, which will then be verb-phrasal.
> (1984:60)

[36] The questionnaire sheet is included in Appendix 5-1.

[Notes on the informant test]

(i) For each example, the acceptability is judged according to the following five-choice-scale. Subjects are required to choose one from the following scale of acceptability:

Table 5.1.15

No problem	More or less OK, but sounds a little bit odd	Sounds odd but not sure if acceptable	Sounds odd and probably unacceptable	Sounds odd and completely unacceptable
+2	+1	0	-1	-2

(ii) To capture the approximate degree of acceptability, a score is assigned to each choice, ranging from 2 ('No problem') to minus 2 ('Sounds odd and completely unacceptable'). The following table shows how the acceptability score for the sentence (34b) is calculated:

Table 5.1.16

	(34b) In Ben's picture of her, what did she find?					
A	Score for each choice	+2	+1	0	-1	-2	
B	number of subjects who select	4	7	1	1	2	15 (total number of subjects)
C	A × B	+8	+7	0	-1	-4	10 (total score)
							Average = 0.67 (Average = total score/ total number of subjects)

The score assigned to each choice (Row A in Table 5.1.16) is multiplied by the number of subjects who select the choice (Row B in Table 5.1.16), and the results of the five choices (Row C) are added (total score) and divided by the total number

of subjects (15), resulting in the average score.

Average score = total score / number of subjects (15)

The average score can range between 2.0 (highest acceptability) and -2.0 (lowest acceptability).

(iii) The alternative version of a Reinhart's example (e.g. 34b) is marked with a single quote (e.g. 34b').

(iv) In addition to Reinhart's examples and alternative examples, some dummy sentences are included.

Table 5.1.17 The result of informant test

[Wh-question]

Sentences		Number of subjects and (Score) for each choice					(Average score)
		+2	+1	0	-1	-2	
(34b)	In Ben's picture of her, what did she find?	4 (8)	7 (7)	1 (0)	1 (-1)	2 (-4)	(0.67)
(34b')	In the lucky bag Dad gave you, what did you find?	10 (20)	2 (2)	1 (0)	2 (-2)	0 (0)	1.33
(35b)	In Detroit, who did the gangsters kill?	8 (16)	5 (5)	1 (0)	1 (-1)	0 (0)	1.33
(35b')	In Detroit, who did the gangsters kill with the shotguns they smuggled?	7 (14)	5 (5)	1 (0)	2 (-2)	0 (0)	1.13
(36b)	In front of Felix, who holds a candle?	2 (4)	6 (6)	2 (0)	4 (-4)	1 (-2)	0.27
(36b')	In front of the demo, who is holding the large picture of Nelson Mandela?	7 (14)	4 (4)	2 (0)	0 (0)	0 (0)	1.6
(37b)	With her new boss, why does Rosa argue?	2 (4)	4 (4)	7 (0)	2 (-2)	0 (0)	0.4
(37b')	With your children, why do you usually argue about their exam results?	4 (8)	6 (6)	3 (0)	2 (-2)	0 (0)	0.8

[Yes/No-questions]

Sentence		Number of subjects					Average
		+2	+1	0	-1	-2	
(38b)	In Ben's picture, did Rosa find a scratch?	7 (14)	4 (4)	1 (0)	3 (-3)	0 (0)	1.0
(38b')	In the new version of Windows, did you find any bugs that you haven't come across before?	12 (24)	1 (1)	0 (0)	2 (-2)	0 ()	1.54
(39b)	With her boss, does she argue?	2 (4)	4 (3)	6 (0)	3 (-3)	0 (0)	0.33
(39b')	With her students, does she usually argue about how she marks their essays and exams?	5 (10)	6 (6)	3 (0)	1 (-1)	0 (0)	1.0

[Imperatives]

Sentence		Number of subjects					Average
		+2	+1	0	-1	-2	
(40b)	In my next picture, find a scratch, if you can!	8 (16)	5 (5)	1 (0)	1 (-1)	0 (0)	1.33
(40b')	In this brand-new product, find any defects, if you can!	5 (10)	6 (6)	2 (0)	1 (-1)	1 (-2)	0.87
(41b)	With your boss, stop arguing!	2 (4)	3 (3)	3 (0)	6 (-6)	2 (-4)	- 0.4
(41b')	With that customers, stop arguing about the quality and defects of our products!	0 (0)	4 (4)	5 (0)	4 (-4)	2 (-4)	- 0.27

We can see in the result of informant test (Table 5.1.17) that only (41b) and its variant (41b') are given minus average scores, which means overall acceptability judgement of them is negative. All other examples are positively judged. Even for the negatively judged (41b), there are two native speakers who judge 'no problem', and three judge 'more or less O.K.'. If Reinhart's claim is correct, all the examples have to be judged either 'completely unacceptable' (score -2) or 'probably unacceptable' (score -1); at least any of them cannot be positively judged. Although the method of this test may be crude and may need further refinements, the voices of the informants clearly disagree with Reinhart's claim in terms of the acceptability of the sentences. Having presented the examples (34b) - (41b), Reinhart claims:

> These facts show that preposed **verb-phrasal** PPs are indeed moved into COMP position, but that **sentential PPs** ... occur in a position preceding the COMP, and since COMP is the leftmost constituent of a sentence, this suggests that they are attached to a higher node than the one dominating COMP.   (ibid. p.69).

Contrary to her claim, however, the overwhelming disagreement in the acceptability judgement leads us to conclude that there is no such a 'fact' to justify to discriminate the syntactic position between initial sentential PP and initial verb-phrasal PP. Referring to Reinhart's examples (34b to 41b), Carden also points out that "there is no independent evidence motivating the structural contrast" between verb-phrasal PP and sentential PP:

> It is true that her proposal for distinct structures for preposed S-Adverbials and VP-Adverbials is ad hoc in the sense that there is no independent evidence motivating the structural contrast ...
> (Carden 1986:328)

There are also other accounts expressing disagreement in the acceptability judgement of examples presented by Reinhart (Van Hoek 1997, Cornish 1986a).

Reinhart's examples with her judgement do not give supportive evidence for her argument but rather serve to demonstrate Reinhart's idiosyncratic acceptability judgement. They are more likely to support Brown and Yule's argument on the problem of "an extreme version of the constructed-sentence approach":

> The essential problem in an extreme version of the constructed-sentence approach occurs when the resulting sentences are tested only against the linguist's introspection. This can (and occasionally did) lead to a situation in which a linguist claims that the 'data' he is using illustrates acceptable linguistic strings because he says it does, as a result of personal introspection, and regardless of how many voices arise in disagreement.
> (Brown and Yule, 1983:21)

I would argue that the overwhelming disagreement involved in the voices of the informants clearly shows the inadequacy of the extreme version of the constructed-sentence approach, which is a consequence of excluding empirical data form the object of inquiry in the generative approaches, due to the influence of Platonic idealism as discussed in 1-2-2.

5-1-6. Empirical contents of Reinhart's model and its problem

In addition to the significant number of reservations expressed by native speakers as to the grammatical status of Reinhart's examples, another problem of Reinhart's account is that it does not allow for many empirically testable statements. In order for the claim to be empirically sound and testable, she needs to specify clearly which initial PP is **verb-phrasal** and which is **sentential**. The following table summarises a few such statements, but many other cases are not provided for.

(5.1.18)

[1] PPs inherently **sentential**	□ PPs that consist of P and S (*after-, when-, because*-clauses) (p.75) □ *in order to*-phrases (p.60) □ *although* ., (causal) *since* ., *whether or not* ., and *with* phrases (p.60)
[2] PPs either **verb-phrasal** or **sentential**	□ infinitive *to*-phrases (p.60)
[3] PPs inherently **verb-phrasal**	□ instrumental (*with* ...) □ manner (*by* ...) □ locatives (e.g. *in* NP) following verbs *dwell, reside, put* □ *with*-phrase following verb *flirt* (p.60)

As noted before, the C-command model cannot be applied to initial **sentential PPs** (listed in [1]). As Reinhart (1984:75) notes below, adverbial clauses are always classified as **sentential PPs**:

> PPs that consist of P and S (such as after-, when- or because-clauses) are always sentential...
> (1984:75)

This means that the C-command model significantly reduces its scope of application in accounting for empirical data of cataphora. The model can no longer deal with the thirteen cataphora examples in AT involving initial adverbial clauses. Having made the claim above, Reinhart gives a qualification to it as in:

However, in cases where it is clear that these PPs must have originated in a clause inside the VP, no 'forward pronominalisation' is possible, as illustrated in (56).

> (56a) * When **Rosa** finishes school, *she* has promised Ben that she will go to London.

What Reinhart says here is that because the initial adverbial clause (PP) is **verb-phrasal** and C-commanded by the main subject NP (*she*), the NP (*Rosa*) in PP has to be a pronoun if it is coreferential with the main subject (*she*); but (56a) breach this constraint. Contrary to her claim, however, Cornish (1986a:249) states that (56a) is "perfectly grammatical and interpretable" (that is, **Rosa** can be coreferential with *she*).

For PPs which are characterised as either **verb-phrasal** or **sentential** (listed in [2]), Reinhart does not show which of them are initial PP is **verb-phrasal** and which are **sentential**, hence there is no way to apply the C-command model to them.

Only to PPs **inherently verb-phrasal** (listed in [3]), can the C-commands model be applied and tested. They can provide either counter-evidence or supportive evidence for the model. Within this limited set of testable cases, is the C-command model empirically tenable? Corpus data such as [S0FEJ095][37], at least, shows that this is not the case.

> [S0FEJ095] **By** an agreement dated 3rd May 1990 between *Mr C A Johnson* and the company , *Mr Johnson* was granted a right to subscribe for up to 5,000 ordinary shares of 25p each in the company .

---

[37] Taken from BNC sample parsed by Liz Eyes.

Figure (5.1.19) Tree for [S0FEJ095], A counter-example to Reinhart's model (1984)

```
                               S'
              _____/  _____
         PP (COMP)                         S
        /      \                      ____/  \____
       P        NP                  NP            VP
       |       /    \               |          /      \
      By   an agreement dated 3rd May   |      was granted a right to
           1990 between Mr C A          Mr      subscribe for up to 5,000
           Johnson and the company    Johnson   ordinary shares of 25p each
                                                in the company
```
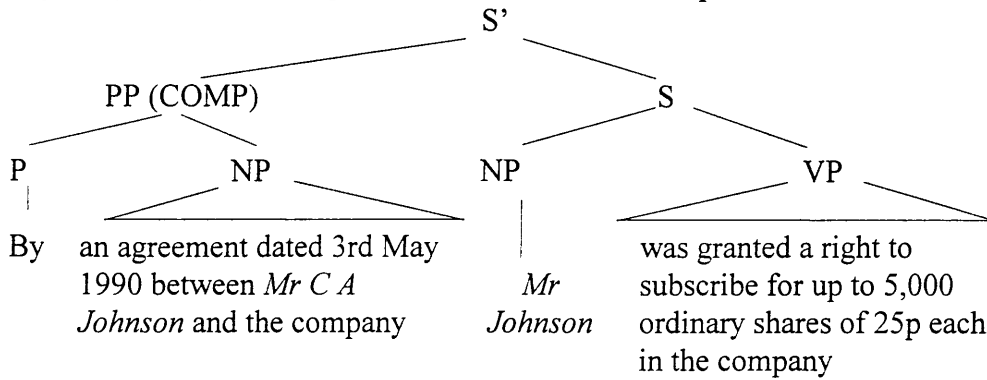
The initial *by* phrase in sentence [S0FEJ095] is manner adverbial phrase which is one of the PPs '**inherently verb-phrasal**' according to the Reinhart's list ([3]). Hence it has to be C-commanded by the main subject *Mr Johnson*. Due to the C-command model, the NP within the initial *by* phrase has to be a pronoun if it is coreferential with the main subject; nevertheless it is not a pronoun (*Mr C A Johnson*), and is capable of being coreferential with *Mr Johnson*.

5-1-7 Conclusion.

In conclusion, it is apparent that Reinhart's C-command coreferential model cannot account for most of the corpus data of cataphora presented in the last chapter. Her constraint cannot claim to be syntactic but highly semantic in nature. It has very limited scope of application to account for empirical data of cataphora (backward anaphora), and its empirical tenability within its domain of application is open to question.

5-2. Backward Pronominalization Rules

In presenting their cataphora data looked at in chapter 4, Carden (1982), Kanzaki (1997) and Van Hoek (1997) argue against an attitude of scepticism towards cataphora such as is represented by the 'Predictability requirement' formulated by Kuno:

> The Predictability Requirement on Backward Pronominalization: Do not pronominalize the left-hand noun phrase unless its referent is determinable (predictable) from the preceding context.   (Kuno, 1975:280)

The genuine cataphora cases (First-mention and Overriding-type) presented in chapter 4 provide conclusive counter-evidence to this 'Predictability requirement', as Carden (1982:385) argues:

> I conclude that the counter-examples survive, that the use of backwards anaphora is not constrained by a predictability requirement, and therefore that we must reject the Forwards-Only Hypothesis.   -

Following the conclusion above, Carden goes on to draw another theoretical

conclusion:

> It follows that sentence-level rules equivalent to the classical rules for backwards anaphora will be needed even in a discourse grammar.

This second conclusion, i.e. the necessity of sentence-level backwards pronominalization rules, will be discussed in this sub-section. I will firstly summarise several points of criticism against generative backwards pronominalization rules in 5-2-1. In 5-2-2, I will explain and discuss the backwards pronominalization rules ("Abstract Model") proposed by Carden (1986). Section 5-2-3 deals with empirical problems of backwards pronominalization rules based on the AT data.

5-2-1. Criticisms against backwards pronominalization rules

It is by no means adequate to represent all the criticisms against the sentential

level backwards pronominalization rules with its extreme/strong form ("Predictability

Requirement" (kuno)) under the label of 'Forward Only Hypothesis' (Carden). There

are, at least, three points of criticism made against the sentence-level backwards

pronominalization rules, which commonly share the following assumption presented

by Ross (1967).

> I will assume that structures underlying sentences like (la) must be converted
> into those that underly (lb) or those that underly (lc) by a transformational rule
> of PRONOMINALIZATION.
>
> (I)(a) After John Adams*i* woke up, John Adams*i* was hungry.
>   (b) After John Adams*i* woke up, he*i* was hungry.
>   (c) After he*i* woke up, John Adams*i* was hungry.

5-2-1-1. Firstly, as illustrated in Ross' passage above, notions of 'pronominalization'

and 'backward pronominalization' necessarily assume the theoretical postulates of the

classical generative framework such as abstract 'deep structures' (Ia), which are

converted to 'surface structures' (Ib and Ic) by the application of 'movement' rules

(such as 'Adverb-preposing' rules) as well as by 'pronominalization' rules. These

theoretical postulates are valid only within the generative framework.   As Bolinger

(1977, Abstract) says, "Pronominalization of one NP by another is a misleading

concept", those who are outside the generative framework cannot accept the notion of

'(backward) pronominalization' itself.

5-2-1-2. Secondly, the 'pronominalization' rules usually assume at first the representation of a sentence at the abstract deep structure level as in (Ia) of Ross' presentation, from which (Ib) and (Ic) are derived. In other words, (Ib) and (Ic) are regarded simply as different manifestations of the same content. This approach is likely to ignore the functional differences in the actual usage between different linguistics forms. In effect, it reduces the issue of the choice of referring expressions to the matter of whether or not transformational rules such as Adverb-preposing are applied to the same abstract deep structure. This kind of approach is hardly acceptable for a theoretical position that claims that different linguistic forms are necessarily associated with the differences in their meaning, as stated by Bolinger

> I suggest that we have been mislead by the concept of 'pronominalization', of an NP in one position somehow controlling an NP in another, and that we ought to be asking instead what is the basis of choice for a noun or a pronoun in any given position, viewing each as an independent decision.
> (Bolinger, 1977:4)

5-2-1-3. Thirdly, one of the main criticisms towards 'backward pronominalization' is that it is impossible to account for pronominal references solely on the basis of sentence-level rules; we need to take account of discourse factors in order to account for cataphora (backwards anaphora). As Kuno (1975) notes, referring to Karttunen (1968):

so-called backward pronominalization involves discourse factors and is not describable in purely syntactic terms.   (Kuno, 1975:284)

I would argue that this criticism towards the generative syntactic backward pronominalization rules is still valid, despite of Carden's argument against the 'Forward Only Hypothesis'.

Carden (1982:361) says:

> While there are various problems in stating a Backwards Pronominalization
> rule precisely, it is clear that some such mechanism will be needed as long as
> we limit the domain of our rules to a single sentence. But what happens when
> we extend our model to permit rules that have an entire discourse as their
> domain? We will then have a rule of Discourse Pronominalization;

By asking "what happens when we extend our model to permit rules that have an
entire discourse as their domain", however, Carden actually shares the same
theoretical position as Kuno, with his criticism towards the sentential approach to pro-
forms, admitting that backwards anaphora cannot be handled solely with a sentence-
level syntactic approach but has to be accounted for through discourse factors. It
should also be noted that the notion of 'First-mention' itself is not a syntactic notion
but a discourse notion. As noted in chapter 4, Carden presents several types of
genuine backwards anaphora. Among them, it is still possible to argue that the
following four types (a)-(d) can be identified without resort to the influence of
discourse factors:

Table 5.2.1: Four types of genuine cataphora (Carden, 1982) identifiable by sentential factors.

(a) First Mention: The antecedent is indefinite NP
e.g. When *she* was five years old, *a child of my acquaintance* announced a theory that she was inhabited by rabbits. (*New York Times*, 6 Nov. 78)
(b) Singular Pronouns Bound by Singular Quantifiers.
e.g. ... there are six legally operated and licensed poker cardrooms . . . As *its* major source of income, *each club* collects a playing fee from the players every half hour. . . (*Social Problems*, 28: 557 (1977))
(c) Plural Pronouns with the Variable Reading.
e.g. Did you know that when *their* wives leave *them*, *two men in five* go bananas? Lanniglan's Rabbi, TV, fall 1977 (collected by T. Dieterich)
(d) Variable-Reading Pronouns Bound by Generics
e.g. No matter how innocent *he* may be in *his* inner soul and in *his* motivations, *the effective mathematician* is likely to be a powerful factor in changing the face of society. (Weiner, *Ex-Prodigy*: 189-190)

The other types of genuine backwards anaphora (the First-Mention cases involving definite NP antecedents and the Overriding types), however, cannot be identified without taking account of discourse factors. We can tell whether the referent of a definite NP is First-mention or not only after checking the preceding text. The overriding types, moreover, necessarily assume more than one competing referent candidate already mentioned in the preceding text.

5-2-2. The backwards pronominalization rules ("Abstract Model") proposed by Carden(1986)

I will firstly explain the pronominalization rules Carden (1986) proposes. The rules are supposed to be the "sentence-level rules equivalent to the classical rules for backwards anaphora", which he claims to be needed "even in a discourse grammar" (Carden, 1982:385). In Carden (1982:379), he says:

> In the structurally-governed type, the proform precedes its antecedent in the same sentence, and obeys a structural condition along the lines proposed by Langacker (1969)/Ross (1969).

As mentioned in the above passage, the pronominalization rules Carden proposes are based originally upon the basic insight of Ross (1967), stating that backwards coreference is possible only when the pronoun is in a subordinate position. Langacker (1969) formalises this condition, often cited as 'Precede and Command' or 'Precede/Command', which Carden (1986:) re-formulates as below[38]:

> (6) The Precede/Command Condition: A pronoun must not both precede and command its antecedent.
>
> (7) Definition: Node X "commands" node Y iff[39] the first "bounding node" B above X dominates Y. .

The definition (7) tells that the domain of 'command' is determined by the definition of "bounding node". Langacker (1969) defines 'S' (clause) as the "bounding node". Due to his definition, in the following sentences,

---

[38] As Carden notes, this formulation is different from Langacker's definition in that Carden's formulation does not include co-ordinate cases, such as:
　　　Penerope cursed *Peter* and slandered *him*.
　　　　　(taken from Langacker, 1969:162)
[39] 'iff' is a standard abbreviation for 'if and only if', indicating the necessary and sufficient condition in logic.

158

(52-1) **His** mother loves **John**.

(cited from Carden, 1986:322)

(52-2) In **his** message to Congress, **Carter** said ...

(cited from the AT [A021 6])

both *His* and *John* in (52-1) command each other since both are dominated by their

first bounding node 'S'; hence they cannot be coreferential with each other, because

the pronoun *His* both precedes and commands *John*, as shown in Figure (5.2.2).

(5.2.2)

```
                    S
           _____/ _____
          /                   \
        NP                     VP
       /  \                   /  \
    SPEC    N               V      NP
     |      |               |      |
    His   mother          loves   John.
```
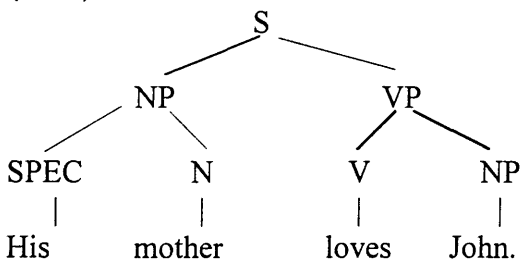
By the same token, since the pronoun *his* both precedes and commands *Carter* in (52-

2), Langacker's 'Precede/Command' rule does not allow those NPs to be coreferential

with each other.

Figure (5.2.3): The tree representation of (52-2)[40].

```
                        S
        _____/|_____
       /                  \      \          \
      PP                    \     NP         VP
     /  \                    \     |        /  \
    P    NP                   \  Carter   said ...
    |   /        \
    In  his message to Congress,
```
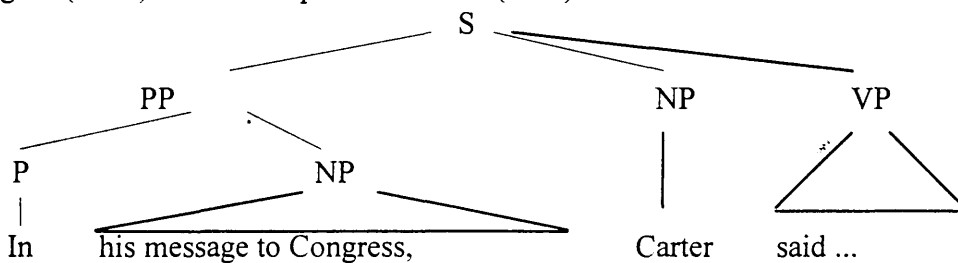
For these particular sentences (52-1) and (52-2), however, coreference can actually be

allowed between the preceding pronouns and the following full NPs. To take account

---

[40] I am following Carden's (1986) notation, which does not use S' nodes but only S
nodes.

of the possibility of coreference in such examples as (52-1) and (52-2), proposals are made that not only S but also NP count as bounding nodes in the definition of 'command' (7)[41]. Due to this modified definition, in (52-1) and (52-2), the first bounding nodes which dominate the 3PPs are not S nodes but NP nodes (*His mother* (52-2) and *his message to Congress* in (52-2)), hence these 3PPs no longer command the supposed antecedent NPs; thus coreference can be allowed between them.

This modified definition of 'Command' (called 'kommand' or 'K-command'), however, still has problems in dealing with what Carden calls 'blocked forward coreference', in which "only backwards coreference is possible" (Carden,1986:320), as in (52-3) and (52-4):

(52-3) Near John, he saw a snake.   (John ≠ he)
(52-4) Near him, John saw a snake.   (John = him)

(5.2.4): The tree representation of (52-3)
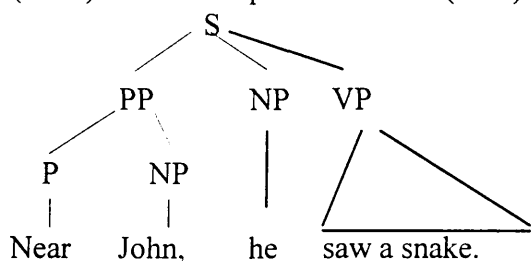
In (52-3), the 'Precede/K-Command' rule assigns coreference between *John* and *he* since coreference is always allowed when the supposed antecedent full NP precedes the pronoun. This prediction, however, does not accord with the actual blocked-forward coreference of (52-3). In (52-4), the first bounding node above the preceding 3PP (NPpro in the Figure (5.2.5)) is the S node which dominate the main subject *John*;

---

[41] See Jackendoff (1972) and Lasnik(1976)).

(5.2.5): The tree representation of (52-4)

```
                S
          /     |     \
        PP      NPa     VP
       /  \      |      /  \
      P   NPpro  |     /    \
      |    |     |    /_____\
     Near him,  John  saw a snake.
```
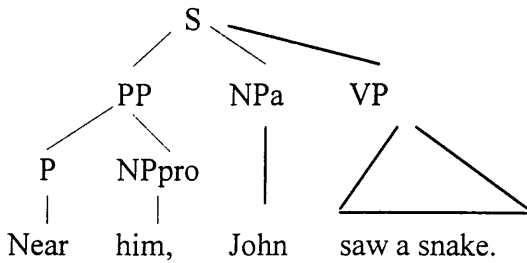
hence the Precede/K-command rule does not allow coreference between *him* and

John, predicting again the opposite of the coreference actually observed.

A possible solution of this 'blocked forward coreference' problem within the

transformational approach can be described as follows:

(i) (52-5) and (52-6) are assumed to be the representations of (52-3) and (52-4)
   (respectively) at the abstract structure level:

   (52-5) He saw a snake near John. (John ≠ he)
   (52-6) John saw a snake near him. (John = him)

(ii) Before the Adverb preposing rule is applied to (52-5) and (52-6), relevant
   coreferential rules are applied to (52-5) and (52-6) at the level of abstract structure,
   resulting in co-indexed (52-7) and (52-8) respectively.

   (52-7) Hei saw a snake near John$j{\neq}i$.
   (52-8) John$j$ saw a snake near him$i{=}j$.

(iii) Then, the Adverb preposing rule is applied to (52-7) and (52-8), deriving (52-9)
   and (52-10) at the surface structure, which accord with the coreference of (52-3)
   and (52-4) actually observed.

   (52-9) Near John$j{\neq}i$, he$i$ saw a snake.   (John ≠ he)
   (52-10) Near him$i{=}j$, John$j$ saw a snake.   (John = him)

This solution, however, invites another problem for the Precede/K-command

rule in dealing with (52-11) and (52-12):

(52-11) After John finished breakfast, he went to school. (John = he)
(52-12) He went to school after John finished breakfast. (He ≠ John)

If we assume that coreference rules are applied to the abstract structure before the Adverb preposing, *he* and *John* in (52-11) have to have the same coreference pattern as *he* and *John* in the unpreposed version (52-12), but in fact (52-11) and (52-12) exhibit different coreferential patterns.

Taking account of the above problems for the Precede/K-command rule, Carden (1986) proposes to assume a reflexive rule in the following way:

(a) The required reflexive rule assigns coreference to reflexive pronouns as well as to certain kind of non- required reflexive pronouns.

(b) The required reflexive rule defines a set of relevant potential antecedents. The potentially coreferential NPs are assumed to be clausemates so that the reflexive rules can be applied to 3PPs and *John* in (52-5) and (52-6) but cannot be applied to *He* and *John* in (52-12).

    (52-5) <u>He</u> saw a snake near <u>John</u>.
    (52-6) <u>John</u> saw a snake near <u>him</u>.
    (52-12) <u>He</u> went to school after <u>John</u> finished breakfast.

The rule needs to exclude typical objects of prepositions from potential antecedents. e.g. *John* in (52-13).

    (52-13) *Bill* talked about <u>John</u> to *himself*.
          (himself = Bill, ≠ John)

(c) The required reflexive rule needs to state the domain of the reflexive:

(c1) to specify that the reflexive morphology is obligatory for coreference with certain kinds of potential antecedents such as direct objects or indirect objects; so that the rule blocks coreference for certain non-reflexive pronouns such as between *John* and *him* in (52-14).

(52-14) John$j$ shot him$i \neq j$.

(c-2) to specify that the reflexive morphology is not obligatory but optional for coreference with certain kind of objects of prepositions; so that the rule assigns coreference to non-reflexive pronouns such as *John* and *him* in (52-8):

(52-8) John$j$ saw a snake near him$i = j$.

Carden calls this required reflexive rule the "Extended Reflexive" rule (ER).

Once the "Extended Reflexive" rule (ER) is stated, the overall steps of Carden's approach to coreference can be described as below:

(1) To the representation of a sentence at the abstract level, apply the Extended Reflexive(ER) rule to assign coreference or non-coreference to reflexives or non-reflexive pronouns as well as identifying the potential antecedent full NPs.

(2) Apply the Adverb preposing rule to the representation of sentence at the abstract level, deriving the representation of sentence at the surface level.

(3) To the representation of the sentence at the surface level, apply the Precede/K-command rule to assign coreference or non-coreference to NPs which have not been assigned coreference.

Carden calls this approach the 'Abstract Model', "since ... it requires that at least one

coreference rule (ER) apply to an abstract structure distinct from surface structure."
(Carden, 1986:324).

The Abstract Model deals with those problematic cases for the Precede/K-command rule, (52-3), (52-4), and (52-11), in the following ways:

For (52-3): Near John, he saw a snake.   (John ≠ he)

Assuming (52-3a) as the representation of (52-3) at the abstract level:

(52-3a) He saw a snake near John.

ER is applied to (52-3a), resulting (52-3b):

(52-3b) Hei saw a snake near John$j≠i$.

The Adverb Preposing rule is applied to (52-3b), resulting (52-3c) at the surface level, which accords with the coreference pattern of (52-3).

(52-3c) Near John$j≠i$, he$i$ saw a snake.

For (52-4): Near him, John saw a snake.   (John = him)

Assuming (52-4a) as the representation of (52-4) at the abstract level:

(52-4a) John saw a snake near him.

ER is applied to (52-4a), resulting in (52-4b):

(52-4b) John$i$ saw a snake near him$i$.

The Adverb Preposing rule is applied to (52-4b), resulting (52-4c) at the surface level, which accords with the coreference pattern of (52-4).

(52-4c) Near him$i$, John$i$ saw a snake.

For (52-11): After John finished breakfast, he went to school. (John = he)

Assuming (52-11a) as the representation of (52-11) at the abstract level:

(52-11a) He went to school after John finished breakfast.

ER is applied to (52-11a), resulting in (52-11b) without assigning coreference, since *He* and *John* are not clausemates.

(52-11b) He went to school after John finished breakfast.

The Adverb Preposing rule is applied to (52-11b), resulting (52-11c) at the surface level:

(52-11c) After John finished breakfast, he went to school.

Since coreference has not been assigned in (52-11c), the Precede/K-command rule is applied to (52-11c), resulting in (52-11d) , which accords with the coreference pattern of (52-11).

(52-11d) After John*j* finished breakfast, he*j* went to school.

5-2-3. Problems to Carden's Abstract model

<u>5-2-3-1. Possibility of non-coreference.</u>

What is neglected in Carden's account is that the non-reflexive pronouns in the following examples cited by Carden can be not only coreferential with but also non-coreferential with *John*.

(52-4)   Near him, John saw a snake.
(52-6)   John saw a snake near him.
(52-11)   After John finished breakfast, he went to school.
(52-12)   He went to school after John finished breakfast.

For instance, if (52-4) is included in the context of (52-16):

(52-16) John was sitting near the baby Peter and watching over him. Near **him**, John saw a snake.

the 3PP *him* in the second sentence is more likely to be coreferential with *the baby Peter* rather than *John*, though the ambiguity between 'John' and 'Peter' still remains. In the similar way, if (52-6) is given a preceding context as in (52-17):

(52-17) John was watching over the baby Peter, who was trying to catch a butterfly. John saw a snake near **him**.

The 3PP *him* is in the second sentence is more likely to be coreferential with *the baby Peter* in the previous sentence rather than with the subject NP *John*. The same point can be made about (52-11) and (52-12) when they have a preceding context as in (52-18) and (52-19) respectively:

(52-18) Peter looked after his father John while John was ill. Peter had to wake up earlier than usual to prepare John's breakfast. After John finished breakfast, **he** went to school.

(52-19) Peter looked after his father John while John was ill. Peter had to wake up earlier than usual to prepare John's breakfast. **He** went to school after John finished breakfast.

Thus if we expand the environment of observation from sentence to discourse, what Carden calls 'blocked forward coreference' can exhibit itself as an ordinary forward reference. No sentence level rules such as the Precede/K-command or the Extended Reflexive rule can assign coreference to the non-reflexive pronouns which are **not** coreferential with *John* in the examples above. This may clearly show that it is not adequate to account for pronominal reference solely in terms of sentence level rules.

5-2-3-2. Different coreference patterns between reflexive and non-reflexive pronouns

Another problem of Carden's Abstract model is that the different coreference patterns between reflexive and non-reflexive pronouns are neglected in Carden's account. Carden (1986:324) states that the reflexive version is merely optional to the non-reflexive version (9b) below:

(9) b. John saw a snake near him. (John = him)

... we will need to state the rule so that reflexive morphology is optional on OPs (objects of preposition) like that in (9b), ...

Carden's example (9b), numbered (52-6) in this section, however, exhibits a different coreference pattern from the refexivised version (52-15):

(52-6) John saw a snake near him.
    (John = him) or (John ≠ him)
(52-15) John saw a snake near himself.
    (John = himself) but unlikely to be (John ≠ himself)

The difference can be recognised clearly when they are observed in the expanded environment, as in (52-16) and (52-20):

(52-16) John was sitting near the baby Peter and watching over him. Near **him**, John saw a snake.

(52-20) John was sitting near the baby Peter and watching over him. Near **himself**, John saw a snake.

Although the non-reflexive pronoun reference in (52-16) still remains ambiguous between 'John' and 'Peter', the reflexive pronoun reference in (52-20) is unlikely to be made to 'Peter'. The same point can be made to (52-17) and (52-17):

(52-17) John was watching over the baby Peter, who was trying to catch a butterfly. John saw a snake near **him**.

(52-21) John was watching over the baby Peter, who was trying to catch a butterfly. John saw a snake near **himself**.

To state that "reflexive morphology is optional" for theses cases means to regard both the reflexive versions and the non-reflexive versions as identical in terms of coreference, failing to capture the different coreference patterns between them.

Even for those cases appearing to exhibit the normal initial-ADV cataphora pattern, the AT data show that they are not necessarily judged by the native speakers as cataphoric but sometimes judged as anaphoric. The following are instances where the 'structural condition' invites an anaphoric reading (with the preceding context)[42]:

[A006 5] A trial court judge said there was insufficient evidence to deport *<REF=1 him*, but an appeals court reversed the decision - prompting *(1 Fedorenko 1)* to file a Supreme Court appeal recently.

[A006 6] In asking the Supreme Court to be permitted to retain *<REF=1 his* citizenship, *(1 Fedorenko 1)* said <REF=1 he had 30 years of good conduct in the United States.

[A022 16] Over the years , *(1 DeLury 1)* gained a reputation on both sides of the bargaining table as {{1 an effective labor leader and an innovative negotiator 1}} .

[A022 17] When *<REF=1 he* retired , *(1 DeLury 1)* was succeeded by the union's secretary-treasurer , Edward Ostrowski, 56 , who said of <REF=1 his death : "We have lost (1 a major figure in our lives 1) "

As discussed in chapter 4, there are good reasons for both the anaphoric reading and cataphoric reading: the anaphoric reading can be supported by the fact that the intended referent is predictable because it is the only candidate referent or the previous focus; for the cataphoric reading, a reader can argue that if the writers had not intended a cataphoric reading, they could have used a pronoun as the main subject rather than a full NP.

In some cases such as those below (3PP *he* in [A31 40], *her* in [A45 70]), the analyst of data cannot even decide whether it is cataphoric or anaphoric.

---

[42] The anaphorically judged cases found in the AT are listed with the preceding context in Appendix 5-2-3.

[A031 39] But *(37 Powell 37)*, asked about that Friday during an informal dialogue with reporters, said " *<REF=37 I* can't deal with that in any specific terms, "other than to refer questioners to (33 Carter 33)'s news conference statement.

[A031 40] Asked how *?<>REF=37 he* would characterize the recent developments in efforts to free the hostages, who were in their 104th day of captivity, *(37 Powell 37)* repeated (33 Carter 33)'s statement in an interview earlier in the day with a group of magazine editors.

[A045 68] *(188 The girl 188)* has been the subject of five successive days of searches by volunteers and police using tracking dogs and helicopters.

[A045 69] On Wednesday, more than 60 volunteers and police searched the area between *<REF=188 her* house and the school, two miles away , but found no evidence of *<REF=188 her* whereabouts.

[A045 70] On the day of *?<>REF=188 her* disappearance, *(188 Cara 188)* overslept and was late for school, <REF=188 her mother said.

What happened here is that the readers/analysts could not decide which NP is the antecedent of the pronoun. This indeterminacy of identifying the antecedent as well as variability of directional judgement (anaphora/cataphora) according to the analysts seems to pose a fundamental question for the traditional notions of 'anaphora' or 'antecedent'. At least, it seems to require us to re-define these notions when we analyse texts from readers perspective. I will come back to this point later in chapter 7. In any case, these AT examples tell us that what Carden calls the 'structurally-governed type' or 'good backward' cases are not necessarily recognised as backward anaphora but sometimes interpreted as ordinary forward anaphora by the analysts/readers.

5-3. Cataphora and Binding Theory

As Binding Theory has been one of the predominant syntactic theories accounting for pronominal references during 1980's and 1990's, it is difficult to avoid mentioning it in discussing cataphoric reference. Although cataphora (backwards anaphora) was one of the central issues in the generative approach in the 1960's and 1970's, hardly any more recent account deals with cataphora (backwards anaphora) from the viewpoint of Binding Theory (BT). Theoretical reasons for this neglect will be discussed in this section, as well as problematic cases for Binding Theory (BT).

5-3-1. An overview of Binding Theory (BT)

Before looking at the theoretical reasons for the neglect of cataphora in

Binding Theory, I will briefly overview Binding Theory as presented by Chomsky

(1981, 1985, 1986)[43].

Binding Theory (BT) concerns syntactic conditions regulating coreferential

relations between nominal expressions in a sentence. According to Chomsky

(1981:188), BT subdivides "nominal expressions into three basic categories":

(A) Anaphors: reflexive pronouns, reciprocal pronouns, and NP-trace[44].
(B) Pronominals: pronouns other than anaphors.
(C) Referential expressions (R-expressions):
                         full NPs.

Binding Theory (BT) assumes that since reflexive pronouns and reciprocal pronouns

(anaphors) "have no capacity for 'inherent reference'" (Chomsky, 1981:188), they

must have an antecedent within certain syntactic conditions. Full NPs (R-

expressions), on the other hand, are assumed to have independent reference; they do

not need antecedents. BT assumes that "the constraint on the interpretation of

pronouns is the converse of that on anaphors" (Haegeman, 1994:225). It "predicts

---

[43] In developing Binding theory, Chomsky modified the definition of 'C-command'
and 'government' several times (Chomsky 1981, 1985, 1986). As a result, there are
different formulations of Binding theory proposed by Chomsky himself. Since the
main concern here is with the predominant, widely-accepted account of Binding
theory rather than with the individual features of Chomsky's own accounts, the
presentation of the theory in this section is based not only on Chomsky's own
accounts but also on well-informed textbooks such as Haegeman (1994) and
Haegeman and Guéron (1999).

[44] 'Trace' or 'PRO' forms are so-called 'non-overt categories' or 'empty categories'.
They are not commonly observable as empirical phenomena but the theoretical
postulations made by Chomsky and others.

that pronouns and anaphors are in complementary distribution .... If both a pronoun

and an anaphor are possible in a position they have different readings" (Haegeman,

1994:229).

BT defines "one principle for each of these categories" (Chomsky, 1981:188)

as below:

> Binding Theory
> (A) An anaphor is bound in its governing category
> (B) A pronominal is free in its governing category
> (C) An R-expression is free.
>
> Throughout, the binding is A-binding. ... each anaphor, pronominal and R-expression is in an A-position, within the range of constructions I am considering here.
> (Chomsky, 1981:188)

In this definition, Chomsky qualifies the range of the application of BT; BT can be

applied only to those NPs occurring at A-position, i.e. the syntactic position of an NP

as argument, to which a thematic role such as 'agent' or 'patient' can be assigned.[45]

The term 'to be bound' can be defined[46]:

---

[45] "... arguments fall into the following categories:
  (i)   overt anaphors
  (ii) pronominals
  (iii) R-expressions
  (iv) clauses
Non-arguments include other non-NP categories as well as NPs that are "non-referential": impersonal *it*, existential there, perhaps idiom chunks, ...arguments are the elements that bear a θ-role. Thus each argument or its trace is in a θ-position, hence an A-position" (Chomsky, 1981:101)

[46] The original formulation of 'to be bound' and 'to be free' is given below:

> (i) α is X-bound by β if and only if α and β are coindexed, β C-commands α, and β is in an X-position.
> (ii) α is X-free if and only if it is not X-bound
>
> ...
> Case (i) and (ii) ... define "bound" and "free" with "X" replaced by "A" or "A-

X is bound by Y if and only if

(i) Y is at A-position (argument position);
(ii) X is C-commanded by Y;
(iii) X and Y are coindexed.

In other words, X 'is bound by' Y iff X is C-commanded by the coreferential Y which is at an argument position (A-position). The term 'to be free' simply means 'not to be bound'. The three principles can be re-stated as follows:

[Principle A] An anaphor (reflexive or reciprocal pronoun) is C-commanded by the antecedent (at A-position) within the 'governing category' of the anaphor.

[Principle B] A pronoun (other than reflexives or reciprocals) is **not** C-commanded by the antecedent (at A-position) within the 'governing category' of the pronoun.

[Principle C] Full NPs are **not** C-commanded by any coreferential NPs at A-position.

I will not go into detail on the notion of 'governing category'[47] here since it is not relevant for the discussion of cataphora and BT in this section.

The following examples (75-1) to (75-3) serve to show how BT works. In (75-1) below:

" (*A-bar):
          (Chomsky, 1981:185)

As Chomsky says that "throughout, the binding is A-binding" (1988:188), I will refer 'binding' without the prefix 'A-'; hence I will say simply 'to be bound' or 'to be free' instead of 'A-bound' or 'A-free'.

[47] The definition of 'governing category' is given:

β is a governing category for α if and only if β is the maximal category containing α, a governor of α, and a SUBJECT accessible to α
(Chomsky, 1981:220)

For further details, see Chomsky (1981) and Haegeman (1994).

(75-1) John$_i$ hurts himself$_i$.

Principle (A) assigns coreference between the anaphor *himself* and the subject *John*, since within the 'governing category' of *himself* (in this case the entire clause) *John* C-commands *himself*. (The notion of C-command defined by Chomsky will be explained in 5-3-2 below). In (75-2)

(75-2) John$_i$ hurts him$_{j,\neq i}$.

Principle (B) does not allow coreference between the pronominal *him* and the subject *John*, since within the 'governing category' of *him* (the entire clause) *John* C-commands *him*. In (75-3):

(75-3) He$_i$ hurts John$_{j,\neq i}$.

Principle (C) does not allow coreference between the R-expression *John* and the subject *He* since *He* C-commands *John*.

5-3-2. Theoretical reasons for the neglect of cataphora

There are at least two theoretical reasons for the neglect of cataphora in

Binding Theory (BT): one is due to the definition of C-command; the other is related

to the issue whether initial adverbials are at A-position or not. I will look at C-

command first.

5-3-2-1. The definition of C-command in Binding Theory BT

In *Barriers* (1986), Chomsky gives the definition of 'C-command' as follows:

α C-commands β iff α does not dominate β and every γ that dominates α
dominates β.

If every γ that dominates α dominates β, then the γ which first dominates α (namely

γ1) necessarily dominates β.

Tree (53.2.1)
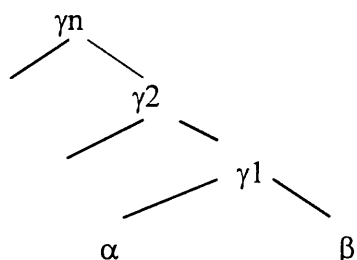
The first γ (γ1) has to be dominated by all the other higher γs (γ2, γ3, ... γn). Hence

there is no need to mention all the other higher γs (γ2, γ3, ... γn) in order to define C-

command; it is enough to mention only the first γ (γ1). It is superfluous to state

"every γ that dominates α" as Chomsky did. Instead of Chomsky's superfluous

formulation above, the notion of C-command can be defined more clearly, as given by

Haegeman (1991):

A node A C-commands a node B iff

(i) A does not dominate B;
(ii) B does not dominate A;
(iii) the first branching node dominating A also dominates B.

or in less negative terms, by Radford (1989:115):

A node C-commands its sisters and their descendants (X is a descendant of Y, if X is dominated by Y)

Two interpretations of C-command have been made depending on how to define the first branching node. The first branching node can be either of any kind or limited to phrasal categories such as S(= IP), NP, VP, PP, AP and S'(=CP). The former is "sometimes referred to as **strict C-command**" (Haegeman, 1994:137), and the latter, the limited type, is called **m-command**. Phrasal categories are termed 'maximal projections' and represented as XP or double-bar nodes in the X-bar representation. In Figure (53.2.2), for example, NP1 *lunch* does not (strictly) **C-command** NP2 *the park* since the first branching node is V' which does not dominate NP2, but NP1 **m-commands** NP2 since the first branching phrasal category is VP which dominates NP2.

Figure (53.2.2)
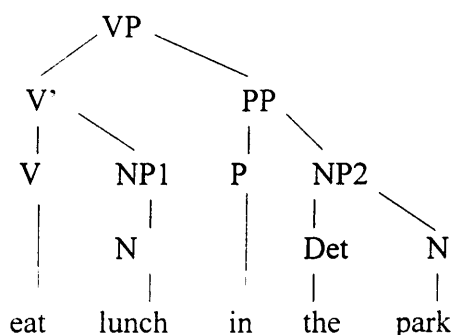
In *Language and Knowledge* (1985), Chomsky limits the first branching node

as "the least maximal projection" (1985:162); hence he defines C-command as equivalent to m-command. In *Barriers* (1986), on the other hand, Chomsky gives the 'strict' definition of C-command as follows[48]:

> α C-commands β iff α does not dominate β and every γ that dominates α dominates β.

> .. It seems that for Binding theory, γ should be taken to be any branching category, ...
> (Chomsky, 1986:8)

Let us examine a typical cataphora construction, the initial-ADV pattern, in terms of both 'strict' C-command and m-command. In the X-bar representation, the main subject position is immediately dominated by S (=IP), and an initial adverbial has to be positioned at some higher-left node to S (=IP), i.e. C or SPEC in Figure (53.2.3).

Figure (53.2.3)
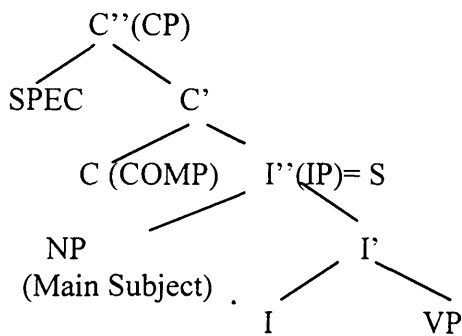
As shown in Figure (53.2.3), the main subject NP cannot 'strictly' C-command any upper-left nodes since the first-branching node dominating the main subject is IP (= S), which cannot dominate any upper-left nodes. The main subject cannot **m-**

---

[48] Notice that the C-command defined by Chomsky is different from the one proposed by Reinhart (discussed in 5-1-1) in that Reinhart's definition allows the main subject to C-command one up-left node (i.e. COMP under C') to which preposed 'verb-phrasal' PPs are attached; though linear order is irrelevant in both definitions.

179

command any upper-left nodes either, since the first phrase category (maximal projection) dominating the main subject is also the first-branching node S (=IP). In any case the main subject NP cannot bind the cataphoric pronoun included in the initial adverbial.

5-3-2-2. The A-bar status of initial adverbials

Let us consider the issue of the status of initial adverbials. Recall that Chomsky qualifies the range of the application of Binding theory (BT) in that BT can be applied only to those NPs occurring at A-position:

> Throughout, the binding is A-binding. ... each anaphor, pronominal and R-expression is in an A-position, within the range of constructions that I am considering here.
> (Chomsky, 1981:188)

Chomsky states that BT can be applied only to those cases where not only binding antecedents but also bound anaphors or pronominals are at A-position. With regard to initial-ADV cataphora, the antecedent is positioned at the main subject, that is, a 'canonical' A-position. The problem is whether an initial adverbial is at A-position or A'(A-bar) position. An initial adverbial is often described as a topicalised structure. In the generative terms, it is a result of a movement to the pre-subject position. Movement to an A-position is called 'A-movement' and movement to an A'-(A-bar) position is 'A'-(A-bar)movement' (Haegeman and Guéron, 1999:218). According to Haegeman and Guéron, "there are a number of arguments against considering topicalization as an instance of A-movement" (1999:218). One such argument is that "the pre-subject position is not associated with any particular grammatical function" (Haegeman and Guéron, ibid.). For instance regarding the following examples:

107   a. To John, I will not say anything.
          b. Clever, she is not.
          c. On that subject, I will consult with John.
          d. Before lunch, I talked to John.
                  (taken from Haegeman and Guéron, ibid.)

Haegeman and Guéron say:

> ... in (107a) we prepose an indirect object, in (107b) we prepose a predicate, in (107c) we prepose an adjunct which modifies the VP, a VP-adjunct, and in (107d) we prepose an adjunct which modifies the clause, a sentential adjunct ...

Since "the pre-subject position is not associated with any particular grammatical function" and that not only NPs but also PPs and APs can occupy the pre-subject position, Haegeman and Guéron conclude that "topicalization is A'(A-bar)-movement" (1999:226). This means that initial adverbials are not in A-position but in A'(A-bar) position; hence they are out of the scope of the application of BT.

5-3-2-3. Problematic cataphora examples for Binding Theory (BT)

Due to the A-bar status of initial adverbials and the fact that an initial adverbial cannot be C-commanded by the main subject under the current definition of C-command, a conclusion can be drawn that Binding theory (BT) as proposed by Chomsky cannot account for cataphora involving initial adverbials. Initial-ADV cataphora is out of the scope of application of BT.

The majority of examples of initial-ADV cataphora data (such as [A033 97] below) involve non-reflexive pronouns in the initial adverbials.

[A033 97] In *his* brilliant career , ***East German luger Dettlef Guenther*** has logged more than 10,000 runs on the frigid , ...

Since initial adverbials cannot be C-commanded by the main subject NPs, these examples, at least, do not provide counter-evidence for Principle (B) of Binding theory (BT).

Problematic cases for Binding theory (BT) are found in those examples in which the main subject has a coreferential reflexive pronoun in the initial adverbial[49]:

[A046 11] In a dissenting opinion for *himself* and Justices William J. Brennan Jr. and Thurgood Marshall, ***Justice Harry A. Blackmun*** objected to "direct aid for ostensibly secular purposes."

---

[49] Further examples are listed in Appendix 5-3.

The main subject *Justice Harry A. Blackmun* , which is coreferential with the

preceding reflexive *himself*, is a First-mention in the text. Hence the 3PP reflexive has

to be linked with the main subject in terms of coreference. Using BT's terms, the

anaphor *himself* must be bound by the antecedent *Justice Harry A. Blackmun* within

the governing category of the anaphor. Due to the definition of C-command made by

Chomsky, however, the main subject cannot C-command and bind the reflexive

himself.

An even more problematic case for BT is what Carden (1986) called the

'Blocked Forward coreference':

(75-3) Near **himself**$_i$, **John**$_i$ saw a policeman.
(75-4) Near **him**$_i$, **John**$_i$ saw a policeman.

The problem of this case is not only that the reflexive *himself* in (75-3) does not have

a binder within its governing category. Recall that Binding theory (BT) assumes that

pronouns and anaphors are in complementary distribution; "the constraint on the

interpretation of pronouns is the converse of that on anaphors" (Haegeman,

1994:225). In other words, pronouns must be free where anaphors (reflexives and

reciprocals) must be bound. "If both a pronoun and an anaphor are possible in a

position they have different readings" (Haegeman, 1994:229). Contrary to this claim,

both the anaphor in (75-3) and the pronoun in (75-4) can be coreferential with the

main subject even though they are in the same syntactic position.

Furthermore, the following two examples taken from CGEL show even cross-sentential reference is possible for the reflexives in the initial adverbials:

No composer enjoyed a better family background than *Mozart*. Like **himself**, both his father and sister were remarkable musicians.

Guerrero's friends made their peace with the gang.
As for **himself**, there was little he could do but await the inevitable bullet in his back.
(CGEL:360)

In these cases, contrary to BT's claim, not only the reflexive form *himself* can be replaced with objective form *him* without affecting the coreferential pattern, but also the anaphor *himself* does not have a binder within the sentence.

Before drawing conclusion about cataphora and Binding theory (BT), I will further explore the problematic examples (not cataphora) for BT in the next section, and discuss Haegeman's counter-argument in the following section (5-3-4).

5-3-3. Further Counter-examples for Binding Theory

Binding theory (BT) predicts that pronouns must be free where reflexives or reciprocals (anaphors) must be bound; "if both a pronoun and an anaphor are possible in a position they have different readings" (Haegeman, 1994:229). The counter-evidence to BT, therefore, usually can be found in a case where both a pronoun and an anaphor can occur at the same syntactic position and where both can have the same coreference pattern.[50] Some such examples are listed in CGEL (p.359) under the title, 'Optional reflexive pronoun'[51].

(1)   In the following examples in which the main subject antecedents C-command the 3PPs, CGEL states that reflexives are optional "in the sense that it may acceptably be replaced with the more usual ordinary objective pronoun." (CGEL:359); hence these examples contradict Principle either (A) or (B):

She's building a wall of Russian BOOKS about *her(self)*.

Mason stepped back, gently closed the door behind *him(self)*, and walked down the corridor.

They left the apartment, pulling the spring lock shut behind *them(selves)*.

---

[50] Horrocks also argues the similar point:

> ... there are counterexamples involving prepositional phrases such as *she pushed the offending object away from her*, where *her* may refer to the same individual as *she* (and *herself* is available as an alternative in this case). (Horrocks, 1987:117)

[51] All examples below are taken from CGEL (pp.359-360).

(2)   In the following example, the antecedent does not C-command the reflexive version; hence it contradicts Principle (A):

Sandra's sister is even taller than *her(self)*.

(3)   CGEL makes the opposite prediction to those of Principle (A) and (B) for the following examples, in which the C-commanding antecedents "require the objective personal pronoun rather than the reflexive pronoun" (CGEL:360):

He looked about *him*.

She liked having her grandchildren around *her*.

She pushed the cart in front of *her*.

They carried some food with *them*.

5-3-4. Haegeman's Counter-argument

In this section, I shall examine in detail the argument made by Haegeman (1994:230-231) to account for a counter example for Binding Theory (BT). Let us consider the following pair of sentences:

(ia) They$i$ saw a snake near them$i$.
(ib) They$i$ saw a snake near themselves$i$.

In (ia), the main subject *They* can be coreferential with the 3PP *them*. The same is true for the main subject *They* and the 3PP themselves in (ib). Since Binding Theory (BT) claims that "if both a pronoun and an anaphor are possible in a position they have different readings" (Haegeman, 1994:229), the pair of sentences is a counter example to BT, as Haegeman admits:

(ia) They$i$ saw a snake near them$i$.
(ib) They$i$ saw a snake near themselves$i$.

If the pronoun *them* in (ia) has the same GC (*governing category) as the reflexive in (ib), then it is unexpected that they may both be coindexed with *they*.
(Haegeman, 1994:230-231 --- *my insertion)

To illustrate in Binding Theory (BT)' terms, if the governing category (GC) for the 3PP *them* is the entire sentence, as indicated by the square brackets in (ia') below:

(ia') [They saw a snake near them].

then, the 3PP must be bound by the main subject *They*.

Hence the Principle (B) of BT tells us that the 3PP must be free in the entire clause, and cannot be coreferential (co-indexed) with *They*.

However, the actual interpretation of (ia) allows the pronoun *them* to be coreferential with the main subject *They*. Hence this is a counter example to BT if the governing category (GC) for the 3PP *them* is counted to be the entire sentence.

In order to deal with this counter-example, Haegeman argues as follows: the governing category (GC) for the 3PP *them* in (ia) is not the entire sentence but a more smaller unit (as shown by brackets in (iia') below) which does not include the main subject *They*, so that the pronoun *them* no longer is bound by *They* within its GC:

(iia') They$i$ saw a snake [near them$i$].

Hence the coreferential relation between the main subject *They* and the pronoun *them* is out of the scope of the application of Binding Theory (BT), and the pronoun *them* can be either coreferential or non-coreferential with *They*; therefore (ia) cannot be counted as a counter-example to BT. I quote Haegeman's whole argument below:

> (ia) They$i$ saw a snake near them$i$.
> (ib) They$i$ saw a snake near themselves$i$.
>
> If the pronoun *them* in (ia) has the same GC as the reflexive in (ib), then it is unexpected that they may both be coindexed with *they*. It is sometimes proposed that in (ia) the PP *near them* is the predicate phrase of a small clause whose subject is non-overt. In (iia) we represent the non-overt subject as PRO. (iia) is roughly analogous to (iib):
>
> (iia) They$i$ saw a snake$j$ [PROj near them$i$].
> (iib) They saw a snake which was near them.
>
> In (iia) PRO, the subject of the small clause, is co-indexed with *a snake* (cf. (iib)). The bracketed small clause is the GC (*governing category) for the pronoun which will duly be free in its GC and may be bound by *they*.
> (Haegeman, 1994:230-231)

Let us examine her argument more closely. In order to claim that the governing category (GC) of the 3PP *them* in (ia) is not the entire sentence but a smaller unit which does not include *They*, Haegeman tries to postulate an abstract entity "non-overt subject" PRO for (ia), as illustrated in (iia):

(iia) They*i* saw a snake PRO near them*i*.

To support her postulation of the non-overt subject PRO, Haegeman claims that (iia) is **roughly analogous** to the relative clause construction (iib).

> In (iia) we represent the non-overt subject as PRO. (iia) is roughly analogous to (iib):
>> (iia) They*i* saw a snake*j* PRO near them*i*.
>> (iib) They saw a snake which was near them.

(iib) is presented to support her postulation of the non-overt subject PRO.

In the relative clause construction (iib), *a snake* is the antecedent of the relative pronoun *which*. Hence, based on her claim that (iia) and (iib) is analogous, Haegeman further claims that *PRO* is co-indexed with *a snake*:

> (iia) They*i* saw a snake*j* PRO*j* near them*i*.
> (iib) They saw a snake which was near them.

In (iia) PRO, the subject of the small clause, is co-indexed with *a snake* (cf. (iib)). Haegeman's insertion "(cf. (iib))" indicates that in order to support her second claim (i.e. co-indexing between *a snake* and *PRO*), she invokes her analogy drawn between (iia) and (iib) again.

Once the abstract entity *PRO* (after the NP *a snake*) is postulated as subject,

the governing category (GC) of the 3PP *them* in (ia) is no longer the entire sentence but the smaller unit indicated by the brackets in (iia):

(iia) They*i* saw a snake*j* [PRO*j* near them*i*].

In (iia), *themi* is not bound by *Theyi* but by *PROj*, which is not co-indexed with *themi*; the 3PP *themi* is 'free' in its governing category (GC). This accords with the claim of Principle B of Binding Theory (BT). Hence if we accept Haegeman's postulation made on (ia) (resulting in (iia)), (ia) is no longer counter-evidence to BT.

The main point of her argument is to present (iib) as being "roughly analogous" to (ia):

(ia) They*i* saw a snake near them*i*.
(iib) They saw a snake which was near them.

Haegeman uses the analogous pair of (ia) and (iib) as supportive evidence to assume an empty category *PRO* which is co-indexed with *a snake*, and to assume the governing category (GC) of *them* as the domain enclosed by brackets in (iia):

(iia) They*i* saw a snake*j* [PRO*j* near them*i*].

If this argument is not an adhoc one and has empirically validity, it must be able to be applied to (ib):

(ib) They*i* saw a snake near themselves*i*.

In the same way as Haegeman does for (ia), we can present the relative clause construction (3b) below, as being analogous to (ib):

(ib) They*i* saw a snake near themselves*i*.
(3b) They*i* saw a snake which was near themselves*i*.

Just as Haegeman does, we must be able to use the analogous pair of (ib) and (3b) as supportive evidence to assume an empty category PRO which is co-indexed with *a snake*, and to assume the governing category (GC) of *themselves* as the domain enclosed by brackets in (3a):

(3a) They*i* saw a snake*j* [PRO*j* near themselves*i*].

For (3a), however, Binding Theory (BT) makes a wrong prediction. Principle (A) states that the anaphor *themselves* must be bound within its GC, hence *themselves* must be co-indexed (coreferential) with PRO*j*, which is co-indexed (coreferential) with a snake*j*. Hence, according to BT's claim, *themselves* must be co-indexed (coreferential) with a snake*j*. This is, however, obviously against the actual coreference pattern.

It is clear that Haegeman's argument to 'save' Binding Theory (BT) from the counter-example (ia) is not valid to (ib). We must, therefore, conclude that her argument is an ad-hoc evasive move, and does not have empirical validity.

5-3-5. Conclusion

Although Chomsky (1981) qualifies the range of the application of Binding

theory (BT) only to those NPs occurring at A-position, BT tends to be defined in such

a way that the three Principles are obligatory for any anaphors, pronominals, and R-

expressions, as in Haegeman (1994:240):

> Principle A
> An anaphor **must be** bound in its governing category.
> Principle B
> A pronoun **must be** free in its governing category.
> Principle C
> An R-expression **must be** free everywhere.

or in the Generalised Binding Theory (Aoun, 1986):

> A. An anaphor **must be** X-bound in its domain.
> B A pronoun **must be** X-free in its domain.
> C. R-expression **must be** free

Since reflexives (anaphors) can appear in the initial adverbials and the problematic

examples above, it is necessary for any kind of BT to qualify clearly the range and the

limitation of its application.

5-4. Summary of Chapter 5

I will summarise the main points in chapter 5:

5-1. C-command model by Reinhart (1976, 1981, 1984)

* A Problematic example for Reinhart's C-command model can be found in a case where a full NP in the initial adverbial co-refers with the following main subject NP (either full NP or pronoun).

    To cope with this problem, Reinhart distinguishes initial verb phrasal adverbials and initial sentential adverbials, and states that the seeming counter-examples are cases of sentential adverbials and out-of-scope of her model. Yet her examples to support her argument could not be supported by the vast majority of the native speakers' acceptability judgements on her examples.

* This overwhelming disagreements in the acceptability of her sentences also indicates the inadequacy of the extreme version of the constructed-sentence approach.

* A misunderstanding of the scope of the application of Reinhart's C-command model can be observed in the account in Kuno (1987), Takami (1985), and Carden (1986). Their misunderstandings seem to be triggered by an inadequate statement made by Reinhart herself (1984:54), which shows a misunderstanding of the scope of the application of her own model.

5-2. Backward pronominalization rules (Carden, 1986)

* What Carden calls 'blocked forward coreference' is 'blocked forward' only when it

is observed in the sentence-level environment. If we expand the environment of observation from sentence to discourse, the 'blocked forward' cases do not always exhibit themselves as 'blocked forward coreference' but as ordinary forward reference.

* The borderline cases (anaphora or cataphora) in the AT tell us that what Carden calls the 'structurally-governed type' or 'good backward' cases are not necessarily recognised as backward anaphora but sometimes interpreted as ordinary forward anaphora by the analysts/readers.

5-3. Binding theory (Chomsky, 1981, 1985, 1986)

* Due to the A-bar status of initial adverbials and the fact that an initial adverbial cannot be C-commanded by the main subject under the current definition of C-command (Chomsky, 1985, 1986), a conclusion can be drawn that Binding theory (BT) as proposed by Chomsky cannot account for cataphora involving initial adverbials. Initial-ADV cataphora is out of the scope of application of BT.
* Yet problematic cases for Binding theory are still found in those examples in which the main subject has a coreferential reflexive pronoun in the initial adverbial, as well as in Carden(1986) 's blocked-forward coreference' cases.

Chapter 6: Cataphora and Cognitive Discourse theories (Mental Representation Theories)

6-0. Introduction: Mental representation theories

With the growth of interdisciplinary interest in human cognition, "traditional and/or formal approaches to anaphora have been re-assessed and enriched" (Wales, 1996 :26) from the point of view of cognition, discourse, and its human agents.

Although there are diverse strands in discourse analysis, cognitive linguistics and pragmatics, many of them have a common view that the cognitive discourse theories need to take account of human memory in modelling human text processing.

This is a rather straightforward emphasis. We do not usually forget the content of a sentence immediately after we process it. A reader has a memory to keep the information obtained from a processed sentence, and use it for interpreting the next sentence, next paragraph, and the rest of the text.

We need, therefore, to assume human memory when we account for the interpretation of a text by a human processor.

The cognitive discourse theories usually assume that the referent of a referring expression has an information gathering point or a storage of information called **'mental representation'** (alternatively 'discourse representation' (Brown and Yule 1983), or 'discourse model' (Webber, 1981)) in a processor's memory.

It should be stressed here that assuming human memory and a mental representation (MR) in modeling human text processing is not the same as assuming the generative mental entity 'competence'. Emmott notes:

> ... 'competence' and cognition are not synonymous. Without needing to accept or even contemplate ideas of innate universal structures, text linguists can speculate about both acquired text-processing strategies and the mind's role, on particular occasions, of making information derived from one part of a text available when another part of the same text is read.
> (Emmott, 1997:10)

The speculation about a mental entity based on commonly observable phenomena is distinct from the speculation about a mental entity accessible only thorough individual intuition. Although human memory and a mental representation itself cannot be observed directly, the mental representation is a part of the linguistic model which has to be tested against commonly observable phenomena. The speculation about 'competence', on the other hand, can only be assessed by individual's intuition which

cannot be commonly accessible and observable by any other individuals.

Although the notion of 'mental representation' is commonly assumed, there is divergence among mental representation theories:

> The introduction of mental representations into linguistic models of reference has led to some confusion over the role of the antecedent. Ariel (1990), for example, uses the term 'antecedent' indiscriminately for both textual antecedent (e.g. pp. 18, 27) and mental representations (e.g. pp. 6, 17, 29, 57, 60).
> (Emmott, 1997:204)

Cognitive discourse analysts such as Brown and Yule (1983) are concerned with on-line (sequential linear) text processing. i.e. how a text is interpreted by a human processor as the text unfolds.

Another main concern in cognitive research is to investigate the relationship between the cognitive status of a discourse entity and various referring expressions. In particular hierarchical approaches such as the 'Given/New taxonomy' (Prince, 1981), 'Accessibility Scale' (Ariel, 1990) and 'Givenness Hierarchy' (Gundel et al, 1993) have been studied substantially, partly because of their universal claims.

Compared with anaphora, however, cataphora (backward anaphora) has hardly been discussed in these theoretical approaches, apart from that of Ariel (1990).

In 6-1, I will discuss Ariel's Accessibility theory and cataphora, examining Ariel's empirical predictions on cataphora against the AT data.

6-1. Cataphora and Accessibility Theory (Ariel, 1990)

In her account on cataphora (backward anaphora), Ariel (1990) makes a few empirical predictions "pending empirical evidence", saying:

> " ... natural data for the type of examples we are discussing have hardly been elicited, and all we can rely upon are our own intuitions." (1990:156).

The AT, as a corpus with discourse annotation, enables us to test empirical claims and predictions made by Ariel.

6-1-1. An overview of Ariel's Accessibility theory

I will briefly overview Ariel's Accessibility theory first. Ariel (1990) assumes that the referents of referring expressions must be mental representations:

> Since it is naive to assume that referring expressions directly refer to physical entities (be they linguistic or other kinds of objects), we must assume that in all cases an addressee looks for antecedents which are themselves mental representations.   (Ariel, 1990:6)

She claims that the choice of a referring expression depends upon "the Accessibility status the mental representation of the referent is assumed to have for the addressee at the current stage of the discourse"(1990:69). The Accessibility status of the referent is marked linguistically according to how easy/difficult the retrieval of the intended referent is. Ariel claims that "high Accessibility Markers imply minimal effort, while Low Accessibility Markers imply greater efforts in recovering the antecedent from the

memory" (1990:53). Her claim is generalised in the Accessibility marking scale

below:

[1] Accessibility Marking Scale (1990:73)   (Table 6.1.1)

	Low Accessibility
a	Full name + modifier
b	Full ('namy') name
c	Long definite description
d	Short definite description
e	Last name
f	First name
g	Distal demonstrative + modifier
h	Proximal demonstrative + modifier
i	Distal demonstrative (+NP)
j	Proximal demonstrative (+NP)
k	Stressed pronoun + gesture
l	Stressed pronoun
m	Unstressed pronoun
n	Cliticized pronoun
o	Extremely High Accessibility Markers (gaps, including pro, PRO and *wh* traces, reflexives, and Agreement).
	High Accessibility

It must be noted that the Accessibility scale above does not account for the case where

a new entity is introduced into the discourse. It accounts for the (initial and

subsequent) reference to an already introduced discourse entity.[52] Ariel presents four

determining factors which "contribute to the assumed Accessibility status of an

antecedent" (1990:28), listed in the table below:

---

[52] " ... the classification into ... three types of Accessibility Markers (High, Intermediate, and Low) is an adequate description of unmarked usage in initial retrievals. Thus, when employed to retrieve entities (rather than to introduce them) for the first time, markers can be divided as to unmarked context-types." (Ariel, 1990:69)

Table 6.1.2

Distance:	The distance between the antecedent and the anaphor (relevant to subsequent mentions only).
Competition:	The number of competitors on the role of antecedent.
Saliency:	The antecedent being a salient referent, mainly whether it is a topic or a non-topic.
Unity:	The antecedent being within vs. without the same frame / world / point of view / segment or paragraph as the anaphor.

(Ariel, 1990:28-29)

In what follows in this section, firstly I will discuss Ariel's misunderstanding of Reinhart's C-command model, which Ariel invokes in dealing with cataphora (6-1-2). Secondly, I will examine Ariel's empirical predictions of cataphora in detail against the AT data (6-1-3). I will, then, discuss the problem of the fundamental assumption of her Accessibility theory related to First mention cataphora (6-1-4).

6-1-2. A problem of Ariel' s Interpretation of C-command.

Ariel invokes Reinhart's C-command model in order to account for what she
calls 'highly cohesive units', namely an **initial embedded clause** and a **preposed PP**
(i.e. initial-ADV constructions), within which cataphoric pronoun references tend to
be involved:

> Syntactically, co-reference is allowed provided a pronoun does not C-
> command its antecedent (Reinhart 1981). In unmarked cases of backwards
> anaphora, this principle is fulfilled when the pronoun occurs in a highly
> cohesive unit vis-à-vis the unit in which the antecedent appears (an initial
> embedded clause, a preposed PP). (Ariel, 1990:155)

As discussed in 5-1-6, however, according to Reinhart (1984), initial adverbial
clauses are always **sentential** and out of the scope of the application of the C-
command model:

> PPs that consist of P and S (such as after-, when- or because-clauses) are
> always sentential...   (Reinhart, 1984:75)

Also preposed PPs are not always **verb-phrasal** but can be **sentential** (cf. 5-1-6).
Hence Reinhart's C-command model is not capable of accounting for initial
embedded clauses and preposed PPs (Ariel's 'highly cohesive units').

In addition, the following passage shows Ariel's misinterpretation of
Reinhart's C-command model:

> ... on the reasonable assumption that the sixteen First-mention backwards
> anaphora cases he cites are representative of his larger sample, it is revealing
> that thirteen of the pronouns (81.25 per cent) are in fact C-commanded by
> their antecedents. None C-command their antecedents.
> (p.159)

Contrary to Ariel's observation, among Carden's 16 First-mention examples[53], 6

examples (37.5%) are out of the scope of Reinhart's C-command model since in these

examples neither the pronouns nor the coreferential full NPs C-command each other

(example (10), (11), (14), (15), (19) and (23)). Moreover, 6 other examples (37.5%)

involve initial adverbial clauses which, according to Reinhart (1984), are always

sentential and out of the scope of the application of the C-command model. This

means that Ariel judges some of those out-of-scope cases as instances where

pronouns are "C-commanded by their antecedents".[54]

Rather than invoking Reinhart's C-command model, I would suggest that it is

more adequate to use the terms 'initial embedded clauses' and 'preposed PPs' for

identifying what Ariel calls 'highly cohesive units'.

---

[53] The 16 Carden's First-mention examples can be found in the 20 examples listed in
Appendix 4-1-1 (excluding the first 4 Already-mentioned examples).
[54] As for the rest of 4 examples ((8), (17), (18), and (24)), it is not clear whether they
belong to so-called **verb-phrasal** PPs, which are C-commanded by the antecedents,
or **sentential** PPs.

6-1-3. Ariel's empirical predictions on Already-mentioned cataphora

In her predictions on cataphora, Ariel treats First-mention cataphora and Already-mentioned cataphora differently. She claims that only First-mention cataphora involves a 'highly cohesive unit', i.e. initial-ADV constructions. In addition to initial-ADV patterns, the 1PSP-initial-DS pattern, which involves a structural dependency (cf. 4-2-2), may also be counted as a 'highly cohesive relation'. Ariel's claim is that structural dependency is necessary only for the First-mention cataphora:

> Dependency is only crucial when the antecedent is a New entity. Only in these cases does the addressee actually rely on the antecedent in order to decipher the identity of the pronoun. Thus, when the antecedent is newly introduced into the discourse ... a speaker has to make sure that the pronoun clause be dependent on the antecedent clause so that the interpretation of the pronoun can rely on material from the independent clause.
> (1990:157)

With regard to Already-mentioned cataphora, on the other hand, Ariel questions whether it can really be counted as cataphora:

> Continuing discourse referents, on the other hand, require quite the opposite context. Some separation between the pronoun and the antecedent domains is required for a Low Accessibility Marker to appear in the same sentence containing a previous reference to the same entity. As a matter of fact, such 'backwards anaphora' cases should probably not count as backwards anaphora at all. (1990:159)
> ... when the entities form part of the discourse already, as is the case with McCray's and Bosch's examples, dependency is not needed at all. What determines whether a so-called backwards anaphora is acceptable or not is not different from the usual Accessibility considerations. Distance and low cohesion (as diagnosed by Bosch 1983) are, therefore, the determining factors. (1990:158)

According to Ariel, the necessary conditions for Already-mentioned cataphora is not a 'highly cohesive unit' (structural dependency) but 'low cohesion' as well as distance

("Some separation between the pronoun and the antecedent domains").

When she talks about Already-mentioned backwards anaphora, she refers to the

invented examples given by McCray (1980) and Bosch (1983):

> *She* was told that if she wanted to get anywhere in this dog-eat-dog world,
> *Mary* was going to have to start stepping on some people.
> (McCray, 1980)
>
> *He* lied to me, and *John* was my friend.
> (Bosch, 1983)

In both examples above, the preceding 3PPs occur as the main subject and the

coreferential full NPs occur later in the sentences. Four instances of this kind of

coreference are found in the AT:

[A016 34] *He* called again and Police Officer David Cheshire went to *Dillard*'s home.

[A016 48] *He* leaves a sister, Hazel Boulais, formerly of Holyoke, who lived with *Corriden* for the past two years.

[A038 126] *He* no sooner had finished when Sen. Gerard Conley of Portland, Democratic floor leader in the Senate, took over *Pachios*' seat with his response to the Republicans.

[A052 3] *It* was closed last year after studies showed brick facade panels were insecurely attached to *the building*'s walls.

These pronouns, however, are hardly to be analysed as cataphora when they are

observed in the larger context:

[A016 31] The police dispatcher advised him to "just go take a couple of aspirin ," **(69 Dillard 69)** said.
[A016 32] "So <REF=69 I did that.
[A016 33] (But) after <REF=69 I lay down, the music jive started up in <REF=69 my ear again."
[A016 34] **<REF=69 He** called again and Police Officer David Cheshire went to (69 Dillard 69)'s home.

[A016 40] **(75 Dr. Thomas F. Corriden of Northampton 75)**, senior medical
   examiner for Hampshire County, died Monday at <REF=75 his home after
   an illness of several months.
   ...
[A016 47] **(75 Corriden 75)**'s wife, Ann, died two years ago.
[A016 48] <REF=75 He leaves a sister. Hazel Boulais, formerly of Holyoke, who
   lived with (75 Corriden 75) for the past two years.

[A038 125] The Republicans wound down their presentation and **(120 Pachios 120)**
   walked into the room, sat down in the still-warm chair and held court
   briefly.
[A038 126] <REF=120 He no sooner had finished when Sen. Gerard Conley of
   Portland, Democratic floor leader in the Senate, took over (120 Pachios
   120)' seat with his response to the Republicans.

[A052 2] The state wants to sue the estate of the late architect Edward Durrell Stone
   and his New York firm, which designed **(11 the 28-story library tower 11)**.
[A052 3] <REF=11 It was closed last year after studies showed brick facade panels
   were insecurely attached to (11 the building 11)'s walls.

As shown above, the main subject 3PPs in these examples are marked as anaphoric

('<') by the AT analysts. These 3PPs are analysed as coreferential not with the

following NPs (i.e. cataphoric) but with the preceding NPs (i.e. anaphoric). So far as

this type of coreferential pattern is concerned, Ariel's sceptical view on Already-

mentioned cataphora has certain empirical support.[55]

When we observe the cataphorically marked pronouns in the AT, however, it

becomes clear that Ariel's claim (i.e. Distance and low cohesion are the determining

factors for the acceptability of Already-mentioned cataphora) receives little empirical

---

[55] Wales, however, points out that "with some subordinators like *if* the 3PP certainly
can occur in a fronted main clause" (Wales, 1996:38), as in:

   (21) a. *She* might be mistaken for a character in her own novel if *City hack
   Audi Bevan* were not real.

support.

Firstly, with regard to 'cohesion', not only First-mention cases but the majority of Already-mentioned cases also involve what Ariel calls 'highly cohesive units', i.e. structural dependency. As shown in Table (6.1.4) below, 77 % of the Already-mentioned cataphora examples involve structural dependency (including both the initial-ADV pattern and the 1PSP-initial-DS pattern).[56]

Table (6.1.4): Summary of occurrences of Initial-ADV pattern and 1PSP-Initial DS pattern in Already-mentioned cataphora cases

	occurrences	
Initial-ADV pattern	39	43 %
1PSP-Initial DS pattern	29	34 %
Other Types	18	21 %
Total	86	100 %

An example of Already-mentioned Initial-ADV pattern:

[A054 107] (133 The 29-year-old Sittler 133) has been feuding with Toronto
          president Harold Ballard and General Manager Punch Imlach for most of
          the season .
[A054 108] After the Leafs traded >*REF=133 his* best friend, right wing Lanny
          McDonald, *(133 Sittler 133)* resigned his captaincy.

An example of Already-mentioned 1PSP-Initial-DS pattern:

[A046 65] "When Stowe picked up the money he was not picking up the money as a
          bribe or participating in bribery," Nageley said.
[A046 66] "He was picking up the money for **Jenrette**, period.
[A046 67] He was a courier."
[A046 68] ">REF=61 *I* absolutely deny that," (61 *Jenrette* 61) said.

---

[56] The 86 Already-mentioned cataphora in the AT are listed in Appendix 4-2-2.

Secondly with regard to 'distance', though Ariel does not clearly specify the distance which enables us to test her empirical predictions, the following passage includes a specific statement about 'distance':

[27] a *She* was told that if she wanted to get anywhere in this dog-eat-dog world. *Mary* was going to have to start stepping on some people.
...
Compare [27a] above with the following, where ... the Distance between the pronoun and the antecedent is very small:

[28] ??Apparently, *she* was told that *Mary* will have to step over a few bodies in order to make it in this tough world. Would you agree with this grim prediction?

Indeed, it makes little sense that the Accessibility of the discourse entity corresponding to 'Mary' should be low enough at the point where the second reference to her is to be made, so the choice of a Low Accessibility Marker in [28] above is not justified. In [27a], on the other hand, a whole sentence separates between the first and the second references to 'Mary', by which time the Accessibility of the mental entity 'Mary' may have dropped.
(Ariel, 1990:155-156)

In this passage. Ariel says that the accessibility of the discourse entity may drop due to a whole intervening sentence (clause) between the initial and last references. Yet anaphoric pronoun coreferences intervened by single clauses/sentences are by no means rare. such as 3PP *He* in [A037 58]. *him* in [A029 101] or *He* in [A020 110]:

[A037 56] The 2-1 opinion was issued by the First District Court of Appeal in the case of **Arthur Davis Miles.** 23. a former infielder for the Montreal Expos farm club in West Palm Beach.
[A037 57] The Expos ' 1975 first-round draft pick was at a press party in August 1977 after the farm team's division championship.
[A037 58] **He** and teammates playfully dove into the Intracoastal Waterway.

[A029 99] "Meet **Sen. Edward Kennedy.** " they said in large type next to a photograph of the presidential contender.
[A029 100] And they gave the date and time for a " walk down Main Street between the post office and the square. downtown Rochester."
[A029 101] For whatever reason - or combination of reasons - hundreds of people were there to meet **him**.

[A020 108] So when the House opened for business Monday, **Frenzel** demanded a roll call vote on a routine matter - approving the journal of the previous day's proceedings.

[A020 109] And 219 members responded - three more than needed for the House to conduct business.

[A020 110] **He** demanded roll call votes on three non-controversial matters.

In these cases, the Accessibility of the referents of the 3PPs has hardly dropped by the time when the 3PPs (bold-typed) are encountered.

Also consider the anaphoric coreference of 3PP *he* in [A052 83] with *Connelly* in [A052 75]:

[A052 75] The people are dead forever - so why rush. " says **Connelly**.

[A052 76] "The law always protects the living.

[A052 77] But it's worth asking whether companies should be allowed to uproot a person's remains for the sake of progress.

[A052 78] And if they can. should they be allowed to just dump numerous graves into a big pit - or at least rebury them individually with proper markers?"

[A052 79] "You can even get into the question of air rights over graves.

[A052 80] There's one Rhode Island church that didn't want to move a graveyard, so it built a building a foot or two over the grave markers.

[A052 81] You have to get down and go through a crawlspace to see the markers."

[A052 82] "If you believe in the soul rising. shouldn't you preserve the area above the grave?"

[A052 83] Laws to protect the dead are especially needed in New England, **he** says, because there are so many small family graveyards that are vandalized, forgotten or are casualties of industrial development.

In spite of seven intervening sentences. the accessibility of the discourse entity 'Connelly' ([A052 75]) seems to be assumed to be still high enough for the use of 3PP *he* in [A052 83].

Ariel says. in [28] (repeated below). the distance between the pronoun and the following antecedent is "very small". "so the choice of a Low Accessibility Marker in

[28] ... is not justified."(ibid.):

[28] ??Apparently, **she** was told that **Mary** will have to step over ...

There are 3 intervening words between the pronoun *she* and the antecedent *Mary* in

[28]. Yet the Distance between the pronoun and the following antecedent is even

smaller in the Already-mentioned cases below. [A028 50] (between *his* and *Janssen*)

and [A033 43] (between *his* and *Reynolds*):

[A028 48] **Janssen** won three Emmy nominations for his role in "The Fugitive," and
his movie credits included "The Shoes of the Fisherman," "The Green
Berets," "Two Minute Warning" and "Marooned."
[A028 49] He returned to television in the 1970s in "O'Hara, United States Treasury"
and "Harry O."
[A028 50] In addition to *his* wife, **Janssen** is survived by his father, Harold Myer of
Mesa, Ariz., and his mother, Bernice Janssen of Los Angeles.

[A033 42] **Reynolds**, 37, a former Roman Catholic priest, wrote the comment on his
response to a questionnaire distributed last month to all 187 state legislators
last month by United Press International.
[A033 43] In *his* apology Friday, **Reynolds** wrote, " Many people have said many
things to me about welfare.

In her statement (repeated below), Ariel assesses whether the accessibility of a

discourse entity is high or low at the point where the reference to the discourse entity is

made:

.. the Accessibility of the discourse entity corresponding to 'Mary' should be
low enough at the point where the second reference to her is to be made....
the second references to 'Mary', by which time the Accessibility of the mental
entity 'Mary' may have dropped. (ibid.)

Consider the references of the 3PPs typed in bold below:

(6.1.3.1) In 1966, Lauck retired to his Lake Hamilton home near Hot Springs. Since his retirement, **he** had served a five-year stint on the state Racing Commission.

(6.1.3.2) When asked if he would ever return to " Saturday Night Live, " the television show where he first gained popularity, Belushi said he would not. When the audience groaned at his response, **he** replied. "Well maybe I will return - when hell freezes over."

(6.1.3.3) "I think this country offers so much opportunity to a young person. You can be a doctor or a lawyer. In some countries, the best thing is to be a ski racer." Kiesel said Friday. Although he didn't mention any countries by name, **he** clearly was referring to the Russian team which had just overwhelmed the United States for the second day in a row at the XIII Winter Olympics.

The referents of the 3PPs typed in bold in these examples can be identified without any difficulties by five English native speaker informants. This indicates that the Accessibility of the discourse entity corresponding to the 3PPs are high enough at the point where the reference to the 3PPs are made. The examples above, actually, are modified versions of the corpus samples below, which exhibit the initial-ADV cataphora construction (the full NPs typed in bold below are replaced with 3PPs in the above examples):

Actual occurrence of (6.1.3.1)

[A047 65] In 1966, Lauck retired to his Lake Hamilton home near Hot Springs.
[A047 66] Since **his** retirement, **Lauck** had served a five-year stint on the state Racing Commission.

Actual occurrence of (6.1.3.2)

[A017 54] When asked if he would ever return to " Saturday Night Live, " the television show where he first gained popularity, Belushi said he would not.
[A017 55] When the audience groaned at *his* response, **Belushi** replied. " Well maybe I will return - when hell freezes over."

Actual occurrence of (6.1.3.3)

[A034 99] " I think this country offers so much opportunity to a young person.
[A034 100] You can be a doctor or a lawyer.
[A034 101] In some countries, the best thing is to be a ski racer." Kiesel said Friday.
[A034 102] Although **he** didn't mention any countries by name, **Kiesel** clearly was referring to the Russian team which had just overwhelmed the United States

for the second day in a row at the XIII Winter Olympics.

Although this thesis does not focus on the issues of choice of referring expressions, a question to be posed to the Accessibility theory here is that why the writers chose Low Accessibility Markers (full NPs) for these cases where the accessibility of referents are high enough for the use of High Accessibility Markers (3PPs) ?

Notice that the distance between the pronouns and the following antecedents in [A017 55] and [A047 66] (1 intervening word) is smaller than in Ariel's example [28]:

[A017 55] When the audience groaned at *his* response. **Belushi** replied. ...

[A047 66] Since **his** retirement. **Lauck** had served ...

[28] ??Apparently, **she** was told that **Mary** will have to step over ...

Also notice that there are coreferential full NPs (underlined) in the immediately preceding sentences:

[A047 65] In 1966. <u>Lauck</u> retired to his Lake Hamilton home near Hot Springs.
[A047 66] Since **his** retirement. **Lauck** had served ...

[A017 54] ... <u>Belushi</u> said he would not.
[A017 55] When the audience groaned at *his* response. **Belushi** replied. ...

[A034 101] In some countries. the best thing is to be a ski racer." <u>Kiesel</u> said Friday.
[A034 102] Although **he** didn't mention any countries by name. **Kiesel** clearly was ...

It may be difficult to account for the occurrences of the full NP (typed in bold) in terms of distance as Ariel claims below:

> Continuing discourse referents. on the other hand. require quite the opposite context. Some separation between the pronoun and the antecedent domains is required for a Low Accessibility Marker to appear in the same sentence containing a previous reference to the same entity. (1990:159)

It seems to be more likely the case that the preferred construction of cataphora (initial-ADV construction) influences the writers' choice of referring expressions in these examples.

Having examined Ariel's empirical claims against the AT data, it becomes clear that:

(i) not only the majority of First-mention cases but also the majority of Already-mentioned cataphora involve structural dependency (initial-ADV and 1PSP-initial-DS pattern).

(ii) In the AT, there are Already-mentioned initial-ADV cases where the occurrences of full NPs are difficult to be accounted for solely in terms of distance.

Hence Ariel's empirical claim (Distance and low cohesion determine the acceptability of Already-mentioned cataphora) are hardly supported, at least, by the AT data. Although this thesis does not focus on writers' perspective, I would suggest that the preferred constructions of cataphora such as initial-ADV patterns influence the choice of referring expressions in those corpus examples.

6-1-4. Problem of First-Mention Cataphora.

At first glance, First-mention cataphora appears to be problematic for the
Accessibility scale since it is impossible to assume an already-activated, accessible
mental representation for the referent of First-mention cataphoric pronoun when a
reader encounters it. Recall, however, that the Accessibility scale does not account for
cases which introduce new discourse entities but the retrieval of the already
introduced discourse entities. This means that Ariel's Accessibility scale does not
account for First-mention cataphora but Already-mentioned cataphora only. First-
mention cataphora cases are simply beyond the application of Ariel's Accessibility
scale, and they do not provide any supportive evidence for or against the scale.

Yet there is still a problem for Ariel's Accessibility theory related to First-
mention cataphora. As noted before, the fundamental assumption of Ariel's
Accessibility theory is that the referent (antecedent) of a referring expression must be
a mental representation:

> Since it is naive to assume that referring expressions directly refer to physical
> entities (be they linguistic or other kinds of objects), we must assume that in
> all cases an addressee looks for antecedents which are themselves mental
> representations.   (1990:.6)

If, as Ariel claims, "in all cases an addressee looks for antecedents which are
themselves mental representations", how can an addressee find the mental
representation of a brand new entity which s/he never came across before in his/her
life? Under Ariel's claim above, there is no way to account for a brand new entity

introduced by any kind of referring expressions. [57] Consider the following First-mention cataphora cases which appear at the beginning of the texts:

[A033 32] Amid demands for *his* resignation. *Democratic state Rep. Russell J. Reynolds* apologized Friday for using a racial slur in a response to a survey.

[A033 97] In *his* brilliant career, *East German luger Dettlef Guenther* has logged more than 10,000 runs on the frigid. writhing conduits built for the spartan sleds and riders who participate in one of the most elemental of winter sport.

Except well-known people who are quite often reported in mass media, the people appearing in news articles are usually unknown to most of the readers of the articles. Those unknown reported people are usually brand new information for most of the readers.

The problem here is that for a reader of the news article who never come across the referents ('Russell J. Reynolds' or 'Dettlef Guenther') before. it is impossible to assume any mental representations of the referents existing in his/her memory. Hence, contrary to Ariel's claim, the reader is unable to find the "antecedents which are themselves mental representations".

What is needed for Ariel's Accessibility theory to be more successful model is to take account of the mental representation for a brand new discourse entity. Once it is taken account of. however. it will become clear that the use of pronouns is not necessarily constrained by the hierarchy of Ariel's Accessibility scale, as indicated by

---

[57] Here I am not talking about new information in a discourse. i.e. new information which is already known to the addressee but does not appear in the previous discourse ('unused' in Prince (1981)'s term). I am talking about new information which the addressee never has come across before in his/her life ('brand new' in Prince (1981)'s term).

First-mention cataphora.

What follows in this chapter

In the last section, I discussed the theoretical problems related to cataphora (in particular the First-mention and Overriding types) for Ariel's Accessibility theory. These problematic examples also have a certain theoretical impact upon other cognitive discourse theories, which will be discussed in this the rest of this chapter. In 6-2, I will firstly discuss how cataphora can be dealt with in other hierarchical approaches to the cognitive status of discourse entities and referring expressions, as presented by Gundel, Hedberg, and Zacharsky (1993). In 6-3, I will discuss the sceptical view on cataphora expressed Cornish (1996). The theoretical implication of cataphora for the cognitive framework will be discussed more generally in 6-4.
It will be shown that none of these accounts from a cognitive linguistics perspective are able to account for the observed phenomena of cataphora.

6-2 Cataphora and the 'Givenness hierarchy' (Gundel et al, 1993).

In a similar spirit to Ariel's Accessibility scale, Gundel, Hedberg, and Zacharsky (1993) (henceforth Gundel et al) present a hierarchical scale of cognitive status applied to discourse entities corresponding to referring expressions. This scale is called the "Givenness hierarchy", in which "each status on the hierarchy is a necessary and sufficient condition for the appropriate use of a different form or forms." (1993:275), as shown in Table (6.2.1):

Table 6.2.1

Cognitive status of discourse entity	How this cognitive status is necessary or sufficient for the use of various referring expressions	referring expression examples ('N' stands for a 'noun')
In focus "The referent is not only in short-term memory, but is also at the current center of attention."	This status is necessary for: appropriate use of zero and unstressed pronominals. (279)	*it*
Activated "The referent is represented in current short-term memory."	Necessary for: use of all pronominal forms Sufficient for: the demonstrative pronoun *that* stressed personal pronouns. (278)	*that* *this* *this* N

familiar "The addressee is able to uniquely identify the intended referent because he already has a representation of it in memory (in long-term memory if it has not been recently mentioned or perceived, or in short-term memory if it has)."	Necessary for: all personal pronouns and definite demonstratives Sufficient for: use of the demonstrative determiner *that*. (278)	*that* N
uniquely identifiable "The addressee can identify the speaker's intended referent on the basis of the nominal alone."	Necessary for: all definite reference Necessary and sufficient for: use of the definite article *the*.	*the* N
Referential "the addressee not only needs to access an appropriate type-representation, he must either retrieve an existing representation of the speaker's intended referent or construct a new representation by the time the sentence has been processed."	Necessary for: use of all definite expressions Necessary and sufficient for: indefinite *this* in colloquial English. (276)	indefinite *this* N
Type identifiable "The addressee is able to access a representation of the type of object described by the expression."	Necessary for: use of any nominal expression Sufficient for: Use of the indefinite article *a* (276)	*a* N

The hierarchy is ordered "from most restrictive (in focus) to least restrictive (type identifiable)" (p.276), and "each status entails (and is therefore included by) all lower statuses" (p.276):

> For example, an entity which is in focus is necessarily also activated, familiar, uniquely identifiable, referential, and type identifiable. However, not all uniquely identifiable entities are familiar and not all familiar entities are either activated or in focus. (Gundel et al, 1993:276)

For the use of personal pronouns, according to the Givenness hierarchy, the referent

needs to be at the level of Familiar status, defined by Gundel et al as follows:

> FAMILIAR: The addressee is able to uniquely identify the intended referent
> because he already has a representation of it in memory (in long-term memory
> if it has not been recently mentioned or perceived, or in short-term memory if
> it has). This status is necessary for all personal pronouns (P.278)

The Givenness hierarchy claims that for the use of all personal pronouns, the intended

referent must at least have achieved the cognitive status of FAMILIAR in the

assumed addressee's memory. In other words, the assumed addressee needs to have a

mental representation of the referent already in their memory, and needs to "able to

uniquely identify the intended referent".

It is apparent, however, that the Overriding type cataphora provides counter-

evidence to the claim of the Givenness hierarchy, since the referent of Overriding type

cataphora cannot be uniquely identified for a reader. First-mention cataphora, in

particular the First-mention cataphora which introduces a brand new entity, also

provides counter-evidence to the Givenness hierarchy, since its referent simply does

not exist in the assumed addressee's memory. Hence it is clear that both First-mention

and Overriding type cataphora cannot be accounted for by the Givenness hierarchy.

It is possible to argue, however, that what the Givenness hierarchy specifies is

an "appropriate use" of a pronoun, and First-mention or Overriding type cataphora

cannot be counted as an "appropriate use" of a pronoun. Although the expression

"appropriate use" may involve a prescriptive tendency, it seems to have certain

cognitive rationale.

This may be the point at which we need to pay attention to the retrieval cost of pronouns, i.e. how much effort a processor (reader) needs in order to retrieve the referent of a pronoun. It would be natural to assume that in many cases, people use a pronoun as a useful 'economical' 'shorthand' referring expression. A referring expression can be a useful 'economical' 'shorthand' when it requires a processor (reader) to excert least processing effort or little retrieval cost. Hence it would be plausible to assume that an ordinary anaphoric pronoun (used as a highlight-maintenance signal) usually requires of a processor the least retrieval cost, and can be said to be a simple useful 'shorthand' referring expression. Cataphora or deferring completion cases, on the other hand, are not simple useful 'shorthand' referring expressions at least from the reader's point of view. The retrieval cost for cataphora or deferring completion case must be greater than the retrieval cost for ordinary anaphora. This may explain why the majority of pronouns are used as ordinary anaphora. (cf. 3-8) In other words, the retrieval cost of the use of a pronoun may be reflected in the distribution of the use of pronouns. Hence we cannot simply satisfy ourselves by finding examples that cannot be accounted for by the Givennes hierarchy or Accessibility scale. We may also need to pay attention to the retrieval cost of cataphoric use of pronouns, though it is beyond the capability of this thesis.

6-3. Cornish's Account on the 'Antecedentless Anaphors' and Cataphora

In discussing what is called 'antecedentless' anaphora (often called 'exophora'), Cornish states that there are two necessary conditions for the existence of antecedentless anaphora: one is "saliency in memory of the referent" (Cornish, 1996:27), namely that the referent has to be already accessible/recoverable in the addressee's memory; the other is that the predication of the 'anaphoric clause' (the clause or phrase in which the pronoun occurs) meet the condition explained below:

> What is necessary ... is the prior existence in the interlocutor's mental discourse model of a conceptual representation of the referent. ... The other necessary condition for the use of 'antecedentless' pronouns is the nature of the anaphoric predication:... The anaphor's host or governing predicator assigns to it an argument or non-argument status ... and its sense in context assigns to it one or more semantic selection restrictions ... In a wider context, the entire anaphoric clause is interpreted as predicating a property of the gradually emerging representation of the referent at issue.
> (Cornish, 1996:39)

Cornish argues that these necessarily conditions of 'antecedentless anaphora' also can be applied to cataphora:

> What is going on in the operation of 'antecedentless' anaphora ... can also serve as a 'model', as it were, for backwards anaphora (or cataphora), where a provisional 'default' interpretation needs to be accessed in terms of the anaphoric clause or phrase as well as the discourse context obtaining at that point, **BEFORE** the relevant antecedent-trigger actually becomes available in the textual flow.
> (Cornish, 1996:39 --- emphasis by Cornish)

What Cornish claims here is that before a reader encounters the antecedent NP ('antecedent-trigger' in Cornish's term), the mental representation of the referent must

be already activated and accessible in the reader's memory through processing the immediate surrounding co-text of the pronoun (i.e. the clause or phrase in which the pronoun occurs) as well as the whole preceding context (either textual or situational). In other words, the addressee is assumed to have "a provisional 'default' interpretation" of the pronoun "BEFORE the relevant antecedent-trigger actually becomes available". Cornish, here, takes the same position as Kuno's 'Predictability Requirement'[58] (Kuno, 1975) except that according to Cornish, the referent of a pronoun is available to a receiver not only from the **preceding** textual or situational context but also from the **following** immediate surrounding clause/phrase. Both Kuno and Cornish assume that the referent of a pronoun must be available to a receiver **before** s/he encounters the antecedent NP. For instance, consider 3PP *his* in [A016 35]:

[A016 31] The police dispatcher advised him to "just go take a couple of aspirin,"
             Dillard said.
[A016 32] "So I did that.
[A016 33] (But) after I lay down, the music jive started up in my ear again."
[A016 34] He called again and Police Officer David Cheshire went to Dillard's home.
[A016 35] Putting *his* ...

When a reader encounters the 3PP *his* in [A016 35], there are two salient competing referent candidates, namely 'Dillard' and 'Police Officer David Cheshire'. It is unlikely that anybody can confidently resolve the 3PP at this point, though the previously focus entity is 'Dillard' which is the subject of the previous sentence as

---

58

The Predictability Requirement on Backward Pronominalization is "Do not pronominalize the left-hand noun phrase unless its referent is determinable (predictable) from the preceding context." (Kuno, 1975:280)

well as the previously pronominalised entity. When the reader processes the whole

initial adverbial, however:

[A016 34] He called again and Police Officer David Cheshire went to Dillard's home. [A016 35] Putting *his* ear next to Dillard's head, ...

s/he may be quite sure that the referent is very likely not to be 'Dillard' but to be

'Police Officer David Cheshire', since 'Dillard' cannot put his ear next to 'his own'

head. In this case, as Cornish claims, the reader can resolve the pronoun before

encountering the antecedent NP *Cheshire*, which is the main subject of the sentence:

[A016 35] Putting *his* ear next to Dillard's head, ***Cheshire*** heard the music also.

There are, however, counter-examples to Cornish's claim, some of which are

already presented in this chapter. Obvious counter-examples are the Overriding-type

of cataphora, such as below:

> Eddie Mathews sacrificed Mantilla to second, and Hank Aaron was walked intentionally. That brought up Joe Adcock. On Haddix second pitch, Adcock swung and sent a towering drive that disappeared over the right-field wall. Haddix had lost his perfect game, his no-hitter, and the ball game.
>   It should have been a three-run home run, but in *his* joy, ...
> (Phil Pepe, *Home Run*: 1 19)
> --- taken from Carden (1993:364):

In this example, it is unlikely that anybody can resolve the 3PP *his* at the last sentence

(whether the referent is 'Aaron' or 'Adcock') even after the whole 'anaphor clause' *in*

his joy is processed. A reader needs to process the rest of the sentence (shown below)

including the antecedent NP *Aaron* for resolving the 3PP.

It should have been a three-run home run, but in *his* joy, ***Aaron*** cut across the field after touching second base.

As Carden notes, "if we replace the final occurrence of *Aaron* with *he*, there is no way to tell whether Aaron or Adcock is intended" (Carden, 1993:364).

The following four First-mention examples occurring at the beginning of texts are also counter-evidence to Cornish's claim:

(Each example includes the context immediately before the antecedent NP)

[A011 57] Predicting **he** will fare "much better than anyone expects" in Sunday's Maine caucuses, ...

[A033 32] Amid demands for **his** resignation, ...

[A033 97] In **his** brilliant career, ...

[A049 49] A slump in fourth quarter camera sales and continuing marketing problems with **its** instant movie system has dropped ...

For these examples, it is still possible to say that the initial adverbial limits the range within which the referent of the pronoun can be identified, or that the initial adverbial provides the background information for the further specification of the referent. As CGEL notes:

> With respect to grammatical functions ... *I* (initial position) is associated with those adverbials that can readily constitute the ground, theme ... or 'scene setting' for what follows.
> (CGEL:491)

Yet it is impossible to say that the referent of the pronoun has already become accessible and uniquely identifiable solely from the context presented above. It is

impossible for any reader to identify a unique "provisional 'default' interpretation" (Cornish) of the 3PP.   A reader needs to process the rest of the sentence including the antecedent NP to resolve the pronoun. The full sentences of the above examples are listed below:

[A011 57] Predicting *he* will fare "much better than anyone expects" in Sunday's Maine caucuses, ***California Gov. Edmund G. Brown Jr.*** Tuesday compared his Democratic presidential rivals to "two peas in a pod."

[A033 32] Amid demands for *his* resignation, ***Democratic state Rep. Russell J. Reynolds*** apologized Friday for using a racial slur in a response to a survey.

[A033 97] In *his* brilliant career, ***East German luger Dettlef Guenther*** has logged more than 10,000 runs on the frigid, writhing conduits built for the spartan sleds and riders who participate in one of the most elemental of winter sports.

[A049 49] A slump in fourth quarter camera sales and continuing marketing problems with *its* instant movie system has dropped ***Polaroid Corp.***'s 1979 earnings to less than half last year's levels.

It is clear that Cornish's claim (i.e. the referent of cataphoric pronoun must be predictable/accessible to the addressee before the antecedent NP is processed) fails to be supported by the corpus example of cataphora presented above.

6-4. Theoretical Implications of Cataphora for Mental Representation Theories

Cataphora (backwards anaphora) has been largely neglected in the cognitive discourse approaches in general. Some cognitive linguists express clear sceptical view on cataphora, such as Stockwell (1995):

> Incidentally, an on-line treatment of reference processing clearly precludes the notion of cataphora (forward co-reference). This is again a structuralist concept derived by treating the text as an object which can be worked through to and fro by the linguist.
> (Stockwell, 1995:165)

Even for those mental representation (MR) theories which do not express clear scepticism, it is hard to accommodate cataphora in their theoretical framework.

One of the main reasons for this neglect is that current mental representation (MR) theories tend to assume that a personal pronoun (in particular a 3PP) is primarily used as 'a recall cue' or a focus continuity marker, which signals to the listener/reader to maintain the focus of a previously established referent (in MR theories' terms, a mental representation which is already accessible and salient in the listener/reader's memory). For instance, Mckoon, Gerrig, and Green (1996) state that a pronoun "confirms accessibility that already exists":

> A pronoun does not create accessibility for itself--it confirms accessibility that already exists. (Mckoon et al. 1996)

They also treat a pronoun "as a recall cue":

> "Successful resolution of a pronoun requires that some entity be sufficiently accessible in the comprehender's discourse representation to provide a unique match to the pronoun as a recall cue; if there is no such unique entity, pronoun resolution may fail"
> (Greene et al. 1994:513)

The above statement made by Greene et al clearly shows that the scepticism towards cataphora comes form the view that the pronouns are primary used for the reference to the unique entity already available in comprehender's memory. According to Greene et al's position, First mention cataphoric pronoun is impossible.

Cornish defines anaphora as "a signal to continue the existing attention focus already established" (Cornish, 1996:22), and states that a 3PP signals the addressee to maintain the already focused entity:

> The pronoun *he* ... presupposes that its referent is already salient, that is, 'in focus' within this stereotypical situation ... and prompts the addressee to maintain that high level of saliency
> (Cornish, 1996:23)

When we observe the 3PP references in the AT, indeed, the majority of 3PPs function as the focus (highlight) continuity markers, and are hence anaphoric.(cf. 3-8)

Yet the genuine cataphora examples presented in this thesis clearly show that the use of 3PPs, at least in news reportage, is not necessarily limited to the function of focus (highlight) continuity marker.

This thesis does not account for the writer's perspective, but it appears to be that there are two types of use of personal pronouns at least in the news reportage. One is ordinary anaphoric usage, i.e. a pronoun is used as a focus (highlight) continuity marker, "a signal to continue the existing attention focus already established" (Cornish, 1996:22). The other type, the cataphoric use, occurs where the writer indicates readers to suspend the interpretation of the pronoun until the writer specifies the referent later in the discourse, usually within the same sentence, as if prompting:

'Wait for a second, I will supply more information about the referent very shortly'.

The current mental representation (MR) theories have concentrated almost

exclusively on the pronoun as a focus (highlight) continuity marker. The 'wait-a-

second' usage, indicated in the cataphora examples in news reportage, has been

neglected.

There are a few accounts of the functional aspect of the initial-ADV cataphora

construction. Givón makes an observation that initial adverbials (preposed clauses)

tend to occur at the bridging point of two thematic units:

> ... pre-posed ADV-clauses appear more typically at paragraph initial positions,
> i.e. at the point of thematic discontinuity.
> Pre-posed clauses may be viewed as coherence bridges at the major
> thematic junctures. Their bridging capacity is further aided by their cataphoric
> semantic connections to the main clause.
> (Givón, 1993:847)

In other words, Givón states that an initial adverbial clause tends to appear at the

beginning of a thematic unit, that is the point at which a new topic is introduced.

Van Hoek, based on the observation of over 500 examples of cataphora,

suggests that a functional role of cataphora (backwards anaphora) is "to serve as a

way of capturing the reader's attention and introducing a new referent in a relatively

concise way." (Van Hoek, 1997b:73)

> I suggest that one of the functions of backwards anaphora may be the concise
> introduction of new information into a discourse (here meaning written text),
> and that typical backwards anaphora constructions serve to make a referent
> somewhat more accessible (in Chafe's sense) before it is fully introduced with
> a full noun phrase.
> (Van Hoek, 1997a:114)

According to Ariel, Biller-Lappin (1983) also argues that "backwards anaphora is a

conventional strategy to introduce New topics". (Ariel, 1990:158)

Givón, Van Hoek and Biller-Lappin's accounts seem to suggest that the initial-ADV cataphora pattern has an already established, conventional function of introducing new information into the discourse. After examining the actual occurrences of genuine cataphora (First-mention and Overriding-type) and the functional accounts made by Van Hoek and Biller-Lappin, I will argue that current mental representation theories need to take account of the cataphoric, 'wait-a-second' use of 3PP.

Before moving to the next chapter, I will discuss a conceptual problem of current mental representation theories in the next section.

6-5. A Problem of Current Mental representation Theories

One of the common practices of the current mental representation theories is to equate the notion of 'referent' with mental representation in memory. Emmott calls her theoretical position as a 'referent in the mind model' (Emmott, 1997:200) as opposed to what she calls 'referent in the text' notion, on which, according to Emmott, Halliday and Hasan's cohesion theory is based. Brown and Yule makes one of the earliest remarks of this kind in their discussion of the distinction between 'exophora' and 'endophora' (Halliday and Hasan, 1976):

> ... the processor has a mental representation. In the one case he has a mental representation of what is in the world, in the other he has a mental representation of a world created by the discourse. In each case he must look into his mental representation to determine reference.
> (Brown and Yule, 1983:201)

The idea of equating 'referent' with mental representation is clearly stated in the following passages made by Ariel (1990) and Cornish(1996):

> ... we must assume that in all cases an addressee looks for antecedents which are themselves mental representations.   (Ariel, 1990:6)

> In all cases, it is within a conceptual representation in the mind of speaker and addressee that the referent is located and accessed and not in either the co-text or the physical situational context.
> (Cornish, 1996:38)

For a character in fiction, as Emmott says, "identifying the mental representation is the goal of interpreting a referring expression such as a pronoun" (Emmott, 1997:203). The same may be true for any kind of abstract notions.

Let us consider in the case of an actual well-known person, for example, the former South African president Nelson Mandela. Imagine the following example in which the speaker X, walking with the speaker Y, found Nelson Mandela on the street, and said to Y pointing to Mandela:

> X: Look! That's Mandela.
> Y: Where?
> X: Over there. At the corner.

In this case, since the speaker Y could not identify the object pointed by the speaker X, he asked "where ?" seeking further specification of the object of pointing. The speaker Y is searching for someone existing in the immediate context. What the speaker could not identify is not the mental representation of Nelson Mandela but the actual person. When the speaker Y asked "where ?", the mental representation of Nelson Mandela in his memory must be already identified and activated, and is 'on standby' to store the incoming information about the referent. It is reasonable to describe, "Y is searching for the referent", in the ordinary sense of the word "referent"; the 'referent' in this case, however, cannot be a mental representation in the memory.

What is involved in the speaker Y's response is not only (i) **to identify and activate the mental representation of the referent** ('referent' here simply means what/who the speaker refers to with verbal expressions) but also (ii) **to search for the source of input** (incoming information) **about the referent.**

For each receiver's cognition, the actual person Nelson Mandela is the origin of information, from which new information about him is created. Through various sources of input (TV, news papers, or books), new information about Mandela comes into each receiver's cognition, updating the content of the mental representation of Mandela in the receiver's memory. The origin of information can overlap with the source of input when we happen to encounter the actual person on the street. The content associated with the mental representation of Nelson Mandela depends on the receiver's knowledge and experiences of Mandela, and differs from individual to individual.

As it is often pointed out, the identification of the source of input is also mediated, controlled and guided by the information stored in the mental representations in our memory. We need to have "a mental representation of what is in the world" (Brown and Yule, 1983:201) for searching an actual person in the immediate context of an utterance.

Yet a mental representation cannot trigger, update and modify itself. A mental representation is "mental stores of information" (Emmott, 1997:9); it cannot be the source of incoming information about the referent. Our cognitive faculty itself cannot be the source of input that triggers updating the content of the mental representations. The elements that trigger updating and revising the mental representation must exist 'outside' the cognitive faculty. In other words, our cognitive faculty is not an autonomous, self-contained closed-system but a dynamic open-system, which always

updates itself through interaction with the outside world. The external world provides the source of input which triggers updating / revising the open-system. To equate 'referent' with 'mental representation' may result in mixing up a source of information with a storage of information, thereby defining the cognition to be a self-contained closed system.

This may be the point where mental representation theories need to go back to Brown and Yule's original emphasis: it is the speaker/writer who refers. Practically, as Lyons (1977:177) says, it is convenient to say 'an NP refers to somebody or something'. In this thesis, I also often use the verb 'refer' having a linguistic expression as its subject. Yet in discussing theoretical issues of reference, we need to be more careful about the use of the term.

Brown and Yule quote the following Strawson, Searle, and Lyons' passages to emphasise their point:

> "'referring' is not something an expression does; it is something that someone can use an expression to do"
> (Strawson, 1950)
>
> "in the sense in which speakers refer, expressions do not refer any more than they make promises or give orders"
> (Searle, 1979:155).
>
> "it is the speaker who refers (by using some appropriate expression)" (Lyons, 1977:177)

Having quoted the passages above, Brown and Yule assert that the agent of referring must be the speaker/writer:

> ... in discourse analysis, reference is treated as an action on the part of the

speaker/writer.
(Brown and Yule, 1983:28)

Following this view, it is the addresser (speaker/writer) who refers, not the verbal expressions nor the receiver (listener/ reader). An addresser refers to somebody or something with a verbal expression. For a receiver's perspective, 'referent' is nothing more than 'what the addresser refers to'. To avoid the confusion in terminology in the current mental representation theories, I would suggest to define the notion of 'referent' simply and vaguely as 'what the addresser refers to', and not to equate it with a verbal expression nor a mental representation.

Chapter 7: Anaphora or Cataphora ? --- Borderline Cases of Reference-Direction Judgement of Pronoun Reference.

7-0 Introduction

Recently, the viability of the notions of reference-direction (anaphora/cataphora) is questioned by some cognitive discourse theorists who focus on on-line text analysis.

Wales notes:

> "In some approaches to discourse processing, in fact, it is possible to dispense with the notion of anaphora virtually altogether; it is certainly now considerably demoted in significance (see Jones 1994; Stockwell 1995)." (Wales,1996:27)

To explore this 'demoted' view, and to investigate how the notion of reference-direction (anaphora and cataphora) can be accounted for in on-line text analyses, I will closely look at the 'borderline' cases in terms of reference-direction judgement.

It should be noted that for investigating this issue, we need a record of the interpretation of a text such as the AT, which is a corpus with discourse annotation. Although what the AT provides is a record of off-line text analysis, we can still gain insights for the on-line analysis from the AT.

In the AT, we can observe the following kind of borderline cases in terms of reference-direction judgement (anaphoric/cataphoric):

(i) Indeterminate cases (i.e. the analyst could not decide whether the reference was anaphoric or cataphoric) (marked with '?<>')
(ii) Cataphorically marked ('>') cases that can also be analysed as anaphoric.
(iii) Anaphorically marked ('<') cases that can also be analysed as cataphoric.

The first section 7-1 explains the cognitive discourse theorists' view on the notion of

reference-direction (anaphora/cataphora) in contrast with the view presented in formal

textual expressions by Halliday and Hasan.

In 7-2, I will take a preliminary look at the borderline cases (i), (ii), and (iii) above.

I will then discuss several factors relevant to the reference-direction judgement in

each section below:

7-3. The initiation point and completion point of interpretation, and the gap between
them.

7-4. Suspension of interpretation caused by structural dependency.

7-5. Immediate initiation and rapid completion of interpretation of pronouns

In 7-6, I will discuss and summarise the overall picture of reference-direction

judgement observed in the AT.

7-1. The Cognitive Discourse Theorists' View On Reference-Direction (Anaphora/Cataphora)

A main motivation underling the cohesion theory (Halliday and Hasan, 1976) is to capture the objective properties of texts:

> This suggests that there are **objective factors** involves - there must be certain features which are characteristic of texts and not found otherwise ... We shall attempt to identify these, in order to establish what are the properties of texts in English ...
> (Halliday and Hasan, 1976:1 -- my emphasis)

Halliday and Hasan's emphasis on objective, formal properties of texts can be seen in their treatment of the references made to the context of situation (what they call 'exophoric references'):

> Exophoric reference ... does not contribute to the integration of one passage with another so that the two together form part of the same text. Hence it does not contribute directly to cohesion as we have defined it. For this reason we shall take only little account of exophoric reference ...
> (Halliday and Hasan, 1976:37)

Halliday and Hasan's emphasis on the objective 'texture' seems to cause their ambiguous statements on pronoun reference. On one hand, they emphasise the semantic nature of pronoun reference, using the terms 'co-reference' and 'co-interpretation':

> There is a semantic link between the reference item and that which is presupposed; but this does not mean that the two necessarily have the same referent.... Co-reference is one particular form that co-interpretation may take - - where the two items do, in fact, refer to the same thing. But the general concept that lies behind the cohesive relation of reference, ..., is that of co-interpretation.
> (Halliday & Hasan, 1976:314)

Yet on the other hand, they use formal textual expressions such as "a pronoun

pointing back to (or 'referring back to') the presupposed NP (antecedent)",

presumably due to their emphasis on the objective properties of texts:

> This form of presupposition pointing BACK to some previous item, is known
> ANAPHORA.
> (Halliday & Hasan, 1976:14 -- their emphasis)

> Characteristically these instances also tend to form COHESIVE CHAINS,
> sequences in which *it*, for example, refers back to the immediately preceding
> sentence -- but to another *it* in that sentence, and it is necessary to go back
> three, four or more sentences ...(ibid. p.15)

These formal textual expressions such as "pointing back to" are criticised by the on-

line discourse theorists, since these expressions do not accord with the on-line

sequential linear processing of texts by human readers:

> ... if you are reading a novel and encounter a new character introduced on the
> bottom of one page, and you turn over the page and find this individual
> subsequently referred to as *he*, it is equally unlikely that it is necessary for you
> to travel back each time through the anaphoric chain to the original *expression*
> to be able to achieve a reference. As a processing model this must be
> implausible.
> (Brown and Yule, 1983:200-201)

Brown and Yule propose to assume that a reader "establishes a referent in his mental

representation of the discourse and relates subsequent references to that referent back

to his mental representation" (Brown and Yule, 1983:200). Based on this view, they

question the distinction between 'exophora' and 'endophora' made by Halliday and

Hasan:

> If this view is correct, the distinction between endophoric and exophoric co-
> reference becomes much harder to draw. In both cases, we must suppose, the
> processor has a mental representation. In the one case he has a mental
> representation of what is in the world, in the other he has a mental
> representation of a world created by the discourse. In each case he must look
> into his mental representation to determine reference.
> (Brown and Yule, 1983:200-201)

I shall contrast the view presented in Halliday and Hasan's formal textual expressions with Brown and Yule's view in the following Figure 7.1.1:

(Figure 7.1.1)
[The arrows in the figure indicate the direction of reference]

	The view presented in Halliday and Hasan's formal textual expressions	*Brown and Yule's view* *(MR = 'mental representation')*
anaphora	full NP ← Pronoun "A pronoun points back to a full NP"	full NP Pronoun ↓ ↓ a MR of a world created by the discourse
cataphora	Pronoun → full NP "A pronoun points forward to a full NP"	Pronoun full NP ↓ ↓ a MR of a world created by the discourse
exhophora	Pronoun ↓ extra-linguistic referent	Pronoun ↓ a MR of what is in the world

It is clear from the figure that in Brown and Yule's model, there is no difference among anaphora, cataphora, and exophora in that all types of reference are made to mental representations (MRs), within which the referent of a referring expression is assumed to reside, as discussed in the last chapter. Hence some MR theorists try to dispense with the notion of anaphora and cataphora virtually altogether from the point of view of on-line processing. One such view can be found in Stockwell (1995):

> Incidentally, an on-line treatment of reference processing clearly precludes the notion of cataphora (forward co-reference). This is again a structuralist concept derived by treating the text as an object which can be worked through to and fro by the linguist. In the reading process, cataphoric markers will simply be seen as incomplete referring expressions, and will be held in short-term memory until a referent is found for them.
> (Stockwell, 1995:168)

I will discuss Stockwell's view on cataphora at the end of this chapter. In what

follows, I will try to investigate how a reference-direction judgement is actually made,

in particular focusing on the borderline cases in the AT.

7-2. Borderline cases in terms of reference-direction judgement in the AT.

7-2-1. Indeterminate cases.

The AT includes the following instances, (7.1) to (7.10), in which pronouns are judged as indeterminate (marked with **?<>REF**), i.e. where the analysts could not decide whether those pronouns were anaphoric or cataphoric.

(Each example is presented with the preceding context including the previous coreferential full NP. All NPs coreferential with the pronoun are typed in bold.)

Cross-deictic coreference

Examples (7.1) to (7.3) below show the same pattern as cataphoric 1PSP_initial_DS cases presented in 4-1, where a singular 1PP in the initial DS has the antecedent full NP in the following reporting clause, hence they can be analysed as cataphoric.

In all of these examples, the referents of the indeterminate pronouns can be said to be highlighted in the previous discourse due to the fact that all of the 1PSPs in the initial DS have coreferential main subjects in the previous sentences. Hence they can also be analysed as anaphoric.

(7.1) *?<>REF=25 I* in [A004 38]

[A004 37] **(25 Alton 25)** said the case of Hill is "the last one" involving Rocky
   Mountain's extension program, which brought the college about $25,000 in
   revenue during the past year.
[A004 38] "That's less than one percent of our budget, though **?<>REF=25 I** guess
   $25,000 is still $25,000." **(25 Alton 25)** said .

(7.2) *?<>REF=29 I* in [A030 32]

[A030 31] **(29 Leonard 29)** was in the right spot at the right time four years ago and
   regrets the same opportunity might not be available for amateur athletes
   this year.
[A030 32] "**?<>REF=29 I** feel really bad for the athletes who have trained so hard for

the Olympics - but we must accept President Carter 's decision," **(29 Leonard 29)** said recently after accepting an award as co-fighter of the year for 1979, along with WBC heavyweight champion Larry Holmes.

(7.3) *?<>REF=61* my and *?<>REF=61 I* in [A046 57], and *?<>REF=61 I* in [A046 58]

[A046 56] **(61 Jenrette 61)**, D-S.C., denied the allegation by the attorney for John Stowe, a Richmond, Va., businessman who introduced <REF=61 him to FBI agents posing as representatives of Arab sheiks.
[A046 57] "To **?<>REF=61 my** knowledge **?<>REF=61 I** never had $50,000 in **?<>REF=61 my** hand in **?<>REF=61** my lifetime.
[A046 58] **?<>REF=61 I** know **?<>REF=61 I** never had $50,000 in cash.
[A046 59] **?<>REF=61 I** wouldn't know what to do with it," **(61 Jenrette 61)** said Wednesday.

Examples (7.4) and (7.5) below are the same as the 1PSP-initial-DS cataphora

type except that the subject of the reporting clause is not a full NP but a pronoun.[59]

(7.4)   *?<>REF=29 I* in [A030 29]

[A030 28] But without the Olympics, **(29 Leonard 29)** says <REF=29 he probably would have gone to college and chosen a more conventional occupation.
[A030 29] "**?<>REF=29 I** wouldn't be a fighter today without the Olympics," <REF=29 **he** said.

(7.5)   *?<>REF=29 my* in [A030 46]

[A030 45] **(29 Leonard 29)** says the Olympics are responsible.
[A030 46] "Not being a heavyweight, which all along has dominated, the Olympics had a great impact on **?<>REF=29** my life." <REF=29 **he** said.

The referent of a 1PSP in a direct speech (DS) has to be the agent of reporting, usually

mentioned in the subject of the reporting clause. Hence the 1PSP *my* in [A030 46]

must be coreferential with the 3PP *he*, that is the subject of the reporting verb "said".

In this regular coreferential relation, the 3PP *he* can be said to be the antecedent of the

1PSP anaphor *my*, as discussed in 2-10-2-3; hence the reference-direction of the 1PSP

---

[59] They might have been encoded by the same analyst, as both belong to the same text.

my is cataphoric. Yet in the highlight-signal relation, the subject of reporting clause 3PP *he* signals the maintenance of the highlight on 'Leonard' (indexed 29), which is triggered by the full NP *Leonard* in the previous sentence. In this respect, the full NP "Leonard" in the previous sentence can be said to be the antecedent of the reporting subject 3PP *he*; hence the reference-direction of the 1PSP *my* can be also anaphoric.

Example (7.6) below is the same pattern as the cataphoric cases presented in 4-1-(g), where a plural 1PP in the initial DS has a coreferential full NP in the reporting clause, hence can be analysed as cataphoric.

(7.6) *?<>REF=51 we* in [A021 49]

[A021 49]   we

[A021 48] Coach Harry Neale was impressed with his old stomping grounds Tuesday
          night but had mixed reviews about **(51 his Vancouver Canucks 51)**.
[A021 49] "That was one of the worst games **?<>REF=51 we**'ve played since
          ?<>REF=51 we left Vancouver," Neale said after **(51 the Canucks 51)**
          earned a 5-5 tie with the Hartford Whalers in National Hockey League
          action.

In this case, the plural 1PP *?<>REF=51 we* can be analysed as anaphoric since it has a coreferential full NP in the previous sentence.

The indeterminate 2PP reference in (7.7) below exhibits the complexity of a cross-deictic coreference.

(7.7)   *?<>REF=60 you* in [A007 49]

[A007 47] **(60 Boyd 60)** said the Sunset now has the worst record in the Amtrak
system even though the Southern Pacific has a good record on its other
trains.

[A007 48] **<REF=60 He** said the Sunset Limited's poor performance reduced the
Southern Pacific's ontime efficiency for all trains to 60.9 percent in
December? .

[A007 49] "Would **?<>REF=60 you** say, to use baseball terminology, that this puts
the Southern Pacific in the second division?" asked Justice Department
Attorney Andrew M. Wolfe.

[A007 50] **(60 Boyd 60)** replied," **<REF=60** I'd say it would put them in the minor
leagues."

The task of resolving 2PP in a direct speech quotation (DS) is to identify the receiver

(addressee) of the DS. Since the initial DS is interrogative in this case, the resolution

task for the 2PP is to identify the person to whom this interrogation is addressed.

At the end of the sentence [A007 49], it becomes clear that the addresser of

the initial DS is not 'Boyd' but 'Justice Department Attorney Andrew M. Wolfe'. A

reader can infer that the interrogation (in the initial DS) by the Justice Department

Attorney is addressed to 'Boyd', who is highlighted in the previous discourse. Hence

a reader can identify the person to whom this interrogation (in the initial DS) is

addressed before starting to process the next sentence [A007 50], the main subject of

which is the coreferential NP *Boyd*. In this respect, the 2PP can be analysed as

anaphoric.

The 2PP in DS can also be referentially linked with the subject of the

reporting clause "Boyd replied" in the next sentence [A007 50], in that the verb

"replied" indicates an interrogation-reply relation in which the recipient (addressee) of

the interrogation is the one who replies. Hence a reader can identify the receiver

(addressee) of the initial DS in [A007 49] to be the agent of the verb "replied",

namely 'Boyd'. In this sense, the 2PP can also be analysed as cataphoric.

A cross-deictic 3PP coreference has very limited formal clues to resolve it; parallelism is one such clue, as shown in (7.8) below:

(7.8)   *?<>REF=36 it* in [A006 32]

[A006 29] The guards put together a list of 21 grievances, topped by a request for an increase in **(36 the base salary 36)** for corrections officers from $728 to $1,500 per month.
[A006 30] Other items on the list included a call for a 20-year retirement plan and for the hiring of additional officers.
[A006 31] Prison Warden Jerry Griffin attended the meeting and spoke to the officers.
[A006 32] "We're working hard to increase **?<>REF=36 it** to at least $840," he said of **(36 the base salary 36)** .

The lexical repetition of "increase" (*increase in the base salary* in [A006 29] and

increase ?<>REF=36 it) can provides a parallelism effect between the coreferential

NP *the base salary* in [A006 29] the 3PP *it* in [A006 32]. This parallelism can give a

clue to resolve the reference of 3PP *it* before a reader encounters the coreferential NP

"the base salary" after "said of". This is despite the intervening sentence [A006 30],

which contains four singular non-human nouns (list, call, plan, and hiring) which

might conceivably be heads of competing antecedent noun phrases. In this respect, the

3PP reference can be analysed as anaphoric.

Because of a writer's limited control over DS, the writer sometimes needs to

supply/encode extra information to avoid potential ambiguity. The expression "said

of" is often used for this purpose. (cf. 2-4-1) In (7.8), *the base salary* after *said of* in

[A006 32] indicates the narrator's information supply for avoiding the potential

ambiguity of 3PP *it* in the initial DS. Due to this indication, a reader is explicitly

made aware of the referential link between the 3PP *it* and the NP *the base salary* after

said of. Hence the reference of 3PP *it* can also be analysed as cataphoric.

Intra-deictic coreference

The 3PP references (marked ?◇REF) in (7.9) below involve the same pattern as the cataphoric initial-ADV, and hence they can be analysed as cataphoric:

(7.9)   *?<>REF=37 he* in [A031 40] and [A031 41]

[A031 39] But **(37 Powell 37)**, asked about that Friday during an informal dialogue with reporters, said "<REF=37 I can't deal with that in any specific terms," other than to refer questioners to **(33 Carter 33)**'s news conference statement.

[A031 40] Asked how **?<>REF=37** he would characterize the recent developments in efforts to free the hostages, who were in their 104th day of captivity, **(37 Powell 37)** repeated **(33 Carter 33)**'s statement in an interview earlier in the day with a group of magazine editors.

[A031 41] Asked if **?<>REF=33/37 he** was more optimistic Friday than **?<>REF=33/37 he** had been in the previous days, **(33 Carter 33)** said <REF=33 he was "more optimistic now than several weeks ago," **(37 Powell 37)** said.

Because the referent of 3PP *he* in [A031 40] ('Powell') is mentioned with the main subject of the previous sentence [A031 39], it can be said to be highlighted in the preceding discourse. In this respect, the 3PP *he* in [A031 40] can be analysed as anaphoric.

The referent of 3PPs *he* in [A031 41] can be either 'Powell' (indexed 37) or 'Carter' (indexed 33). Since both 'Powell' and 'Carter' are mentioned in the previous sentence [A031 40], the 3PP *he* in [A031 41] can be analysed as anaphoric.

The 3PP *her* in [A045 70] below can be analysed as cataphoric since it exhibits the same pattern as a cataphoric initial-ADV:

(7.10) *?<>REF=188 her* in [A045 70]

[A045 66] **(188 Cara Quinn 188)**, an 11th grader at Joel Barlow High School, has been missing since last Thursday when **<REF=188 she** failed to return home in time for a piano lesson.

[A045 67] "**<REF=188 Her** piano lesson was at 5:30," said **<REF=188 her** mother, Joan Quinn, "and that's something we never missed."

[A045 68] **(188 The girl 188)** has been the subject of five successive days of searches by volunteers and police using tracking dogs and helicopters.

[A045 69] On Wednesday, more than 60 volunteers and police searched the area between **<REF=188 her** house and the school, two miles away, but found no evidence of **<REF=188 her** whereabouts.

[A045 70] On the day of **?<>REF=188 her** disappearance, **(188 Cara 188)** overslept and was late for school, her mother said.

A reader can identify the referent of 3PP *her* as 'Cara' before encountering the following coreferential full NP subject, since 'Cara' is the only candidate for female-highlight in the previous discourse, and is pronominalised twice in the previous sentence. Hence the 3PP also can be analysed as anaphoric.

In all of the indeterminate cases examined above, both anaphoric and cataphoric readings have certain evidence to support them. The majority of the examples (7 out of 10) exhibit the typical cataphora patterns (initial-ADV or 1PSP-initial-DS), hence they can be analysed as cataphoric. They can also be analysed as anaphoric since the referents of the pronouns are previously highlighted entities and a reader is likely to be able to predict them before encountering the following coreferential full NPs.

In addition to the indeterminately marked examples, there are other borderline cases in terms of anaphora/cataphora judgement, which, according to the AT annotators, are already anaphoric or cataphoric. As shown in 4-4, initial-ADV type pronoun references are not necessarily judged as cataphoric but are sometimes judged as anaphoric, as in (7.11) and (7.12):

(7.11) 3PP *his* in [A006 6]

[A006 5] A trial court judge said there was insufficient evidence to deport **<REF=1 him**, but an appeals court reversed the decision - prompting **(1 Fedorenko 1)** to file a Supreme Court appeal recently.
[A006 6] In asking the Supreme Court to be permitted to retain **<REF=1 his** citizenship, **(1 Fedorenko 1)** said <REF=1 he had 30 years of good conduct in the United States.

(7.12)   3PP *he* in [A022 17]:

[A022 16] Over the years , **(1 DeLury 1)** gained a reputation on both sides of the bargaining table as {{1 an effective labor leader and an innovative negotiator 1} .
[A022 17] When **<REF=1 he** retired , **(1 DeLury 1)** was succeeded by the union's secretary-treasurer , Edward Ostrowski, 56 , who said of <REF=1 his death : "We have lost (1 a major figure in our lives 1)"

In the cataphorically marked initial-ADV cases, there are instances where anaphoric reading is also possible, as in (7.13): the 3PP *his* in [A017 55]:

(7.13) 3PP *his* in [A017 55]:

[A017 54] When asked if **>REF=38 he** would ever return to "Saturday Night Live," the television show where **>REF=38 he** first gained popularity, **(38 Belushi 38)** said <REF=38 he would not.
[A017 55] When the audience groaned at **>REF=38 his** response , **(38 Belushi 38)** replied , "Well maybe <REF=38 I will return - when hell freezes over."

'Belushi', the referent of the 3PP *his* in [A017 55], is previously highlighted in that it is the subject of the main clause as well as the only pronominalised entity in the previous sentence (for three times). Hence there is a good reason to analyse this 3PP *his* as anaphoric, as distinct from earlier instances in [A017 54].

The same point can be made about (7.14) below:

(7.14) 3PPs *his* in [A025 36]

[A025 35] Persons who work at educational collaboratives, where handicapped
children are instructed, are entitled to the same collective bargaining rights
as other public school teachers, **(46 a state Labor Relations Commission
official 46)** has ruled .
[A025 36] In >*REF=46 his* ruling, *(46 hearing officer Stuart A. Kaufman 46)*
cleared the way for the Massachusetts Federation of Teachers to establish
unions at the Shore Collaborative, which serves several of Boston 's
northern suburbs, and the South Shore Educational Collaborative ,
stretching from Braintree to Scituate.

The 3PPs *his* in [A025 36] marked as cataphora can also be analysed as anaphora because:

(i) The referent (indexed '46') is mentioned in the main subject of the previous

sentence, hence it can be said to be previously highlighted entity.

(ii) The main clause of the previous sentence, "a state Labor Relations Commission

official has ruled" indicates the agent of 'ruling', who must be the referent of 3PP

his in "In his ruling".

Since both the anaphoric reading and cataphoric reading for these cases are

well supported as discussed in 4-1-2, it can be expected that the reference-direction

judgement for them can vary according to different analysts.   The variability

of reference-direction judgement according to analysts may well be demonstrated by the following AT data. The AT includes two distinct news articles dealing with the same topic: the article starting from [A033 32] to [A033 54], and the article starting from [A035 82] to [A035 101]. Not surprisingly, when the two distinct articles report the same quotation, almost identical reporting clauses appear, as in [A033 43] and [A035 99]:

[A033 43] In his apology Friday, Reynolds wrote, "Many people have said many things to me about welfare.

[A035 99] In his Friday apology, Reynolds wrote, "Many people have said many things to me about welfare.

Apart from the ordering of "Friday" and "apology", the two sentences are identical. Yet what is significant is that the reference-direction of the 3PP *his* in the initial adverbials is judged differently between the two instances. One is cataphoric:

[A033 43] In >*REF=33 his* apology Friday, *(33 Reynolds 33)* wrote, ...

The other is anaphoric:

[A035 99] In <*REF=104 his* Friday apology, *(104 Reynolds 104)* wrote, ...

The 3PP *his* in [A035 99] is marked anaphoric presumably because the preceding discourse mentions that the referent 'Reynolds' "issued a written apology Friday" (as underlined in [A035 83] below).

[A035 83] **(104 State Rep. Russell J. Reynolds, D-West Haven 104)**, <u>issued a written apology Friday</u> for his use of the word "nigger."
        ...
        ...
[A035 99] In <**REF=104 his** Friday apology, **(104 Reynolds 104)** wrote, "Many people have said many things to me about welfare.

Yet it could also be analysed as cataphoric since it exhibits the initial-ADV cataphora pattern. Another example [A033 43] is marked as cataphoric, as it exemplifies the initial-ADV cataphora pattern. This could also be analysed as anaphoric, since it is previously highlighted by the main subject full NP "Reynolds" in the previous sentence [A033 42]:

[A033 42] **(33 Reynolds 33)**, 37, a former Roman Catholic priest, wrote the comment on his response to a questionnaire distributed last month to all 187 state legislators last month by United Press International.

[A033 43] In *>REF=33 his* apology Friday, *(33 Reynolds 33)* wrote, "Many people have said many things to me about welfare.

All the borderline cases (indeterminate cases, anaphorically marked initial_ADV cases, cataphorically marked initial_ADV cases) show that both anaphoric and cataphoric readings have certain supportive evidence. This indicates that reference-direction judgements, in practice, cannot be made in terms of a binary choice (either anaphoric or cataphoric). On the contrary, the judgements have to be made according to, at least, a triple-choice (anaphoric / cataphoric /indeterminate), or can involve a cline (between an anaphoric end and cataphoric end of the scale).

In any case, further corpus data and careful psycholinguistic experiments will be needed to investigate the reference-direction judgement.

In the following sections, I will discuss some of the factors relevant to an account of the reference-direction judgements marked in the AT.

7-3. Initiation point and completion point of interpretation.

At a certain point in the on-line processing of a text, the cognitive status of the referent of a pronoun in a reader's memory cannot be considered as a mirror image of that in a writer's memory. At the time when a writer uses (writes down) a pronoun, the referent of the pronoun must be highly focused in his/her own memory. At the time when a reader encounters the pronoun, on the other hand, this is the point when s/he starts to interpret/resolve the pronoun. In cognitive terms, it is the point when a reader starts to search for, identify, and activate the mental representation of the referent of the pronoun.

In some cases, a reader can identify the referent of a pronoun almost immediately with high degree of certainty. This is the case, for instance, when grammatical constraints operate, such as when a reflexive pronoun is the direct object of a transitive verb, as below:

[A054 70] *More candidates* may commit *themselves* after Tuesday's New Hampshire
    primary .

As soon as s/he encounters the reflexive 3PP *themselves*, a reader will immediately resolve it as being coreferential with the clause subject *More candidates*. The same is true when there is only one focused referent candidate of a 3PP in the discourse, such as 3PP *he* in the first sentence of a text, [A003 14]:

[A003 14] *The vice president of the West German Olympic Committee* said Sunday
    he believes the Summer Olympics cannot be held in Moscow this year.

A reader will immediately interpret the 3PP *he* as being coreferential with the clause

subject *"The vice president of the West German Olympic Committee"*.

Yet there are cases in which the resolution cannot be achieved immediately

with a high degree of certainty. Consider:

[A003 28] The government contended <u>Jacobson</u>, 48, former big-time horse trainer
turned East Side real estate operator, killed <u>Tupper</u> because Miss Cain, **his**
...

When a reader encounters the 3PP, there are two referent candidates for the 3PP *his*,

namely 'Jacobson' and 'Tupper'. At this point, 'Jacobson' is the focused entity of the

sentence as it is mentioned in the main subject and elaborated with an appositive

phrase, while 'Tupper' is the most recently mentioned 3PP reference candidate.

Although 3PP tends to be linked with the focused entity, it is rather difficult to predict

the referent with a high degree of certainty at this point. The uncertainty of resolution

still remains at the end of the appositive phrase, *"his live-in girlfriend of five years"*:

[A003 28] The government contended <u>Jacobson</u>, 48, former big-time horse trainer
turned East Side real estate operator, killed <u>Tupper</u> because Miss Cain, **his**
live-in girlfriend of five years, ...

The referent will not become certain (to be 'Jacobson') until a reader encounters

"Tupper" (typed in bold below) towards the end of the verb phrase:

[A003 28] The government contended <u>Jacobson</u>, 48, former big-time horse trainer
turned East Side real estate operator, killed <u>Tupper</u> because Miss Cain, **his**
live-in girlfriend of five years, moved from his apartment to **Tupper**'s just
down the hall.

Consider also the 3PP *he* (typed bold) in [A059 39]:

[A059 38] Freshman Mike Yastrzemski gave evidence of being a chip off the old
block by blasting a three-run, ninth-inning home run to cap a Florida State

comeback for a 9-8 win over the Miami Hurricanes in college baseball action.

[A059 39] The son of veteran Boston Red Sox star <u>Carl Yastrzemski</u> hit the winning blow as <u>his</u> father watched from the stands, as **he** has ...

There are two referent candidates for the 3PP he in [A059 39], namely 'Freshman Mike Yastrzemski' and his father 'Carl Yastrzemski'. 'Mike Yastrzemski' is mentioned with the main subject, while his father 'Carl Yastrzemski' is mentioned with the subject of the previous adverbial clause. As 'Mike Yastrzemski' is mentioned in the main subject, it can be the sentence topic in [A059 39]. In this respect, 'Mike Yastrzemski' is more likely to be the referent of 3PP rather than 'Carl Yastrzemski'. This initial prediction, however, may be reversed towards the end of the sentence:

[A059 39] The son of veteran Boston Red Sox star Carl Yastrzemski hit the winning blow as his father watched from the stands, as **he** has been during the three-game series that concluded here Sunday.

These anaphorically marked examples tell us that the interpretation of a pronoun, even if it is anaphoric, cannot be always stabilised immediately when a reader encounters the pronoun. When a reader cannot be sure about the referent at the encounter-point, the final resolution will be deferred until s/he can obtain enough information to resolve it from the later input. Based on a survey of psycholinguistics experiments[60], Sanford and Garrod (1989) suggest that achieving pronoun resolution, at least its making the final decision, can be deferred until further evidence is obtained through the processing of later context:

> ... although processes supporting resolution may be initiated on encountering a pronoun, full resolution may depend upon later evidence.
> Indeed, there is some indication that even when unambiguous resolution is

---

[60] For the detail of the experiments, see Sanford and Garrod (1989).

possible. later evidence may be taken into account in making a final assignment.   (Sanford and Garrod, 1989:257)

Hence we can assume a distinction between (i) the point when a reader encounters a pronoun and initiates its interpretation (**initiation point**), and (ii) the point when the reader completes to interpret the pronoun (**completion point**).[61] As Sanford and Garrod note:

> It is a truism that all processes take time, and so it is necessary to discriminate between the onset of processes and the point of their completion (Sanford and Garrod. 1989:239)

We can assume a **gap** between the initiation point and the completion point. In some cases. the gap can be almost nil when a reader resolves a pronoun immediately after s/he encounters it. In other cases. the gap can be extended to the end of the phrase, clause. or sentence in which the pronoun is included.

During the gap between the initiation point and the completion point, the information obtained through processing the co-text of the pronoun needs to be held in some kind of temporary location of memory until the resolution is stabilised. In [A003 28] for instance:

[A003 28] The government contended <u>Jacobson</u>. 48. former big-time horse trainer turned East Side real estate operator. killed <u>Tupper</u> because Miss Cain. **his** live-in girlfriend of five years. moved from his apartment to **Tupper's** just down the hall.

Before encountering "Tupper's" (typed bold). a reader will obtain the following information related to the referent of 3PP:

---

[61] It is a truism that the process of text interpretation always involves uncertainty and it is almost impossible to talk about 100% completion of it.. Yet in this thesis. I will tentatively use the term 'completion point' referring to the time/phase when the interpretation of a pronoun is stabilised.

(i) The referent of 3PP has a five years live-in girlfriend, Miss Cain.

(ii) Miss Cain moved from the referent's apartment to somebody else's apartment.

This information cannot be linked with a discourse entity in longer-term memory (namely the referent of the 3PP) with a high degree of certainty until the 3PP is resolved. The information needs to be held in the temporary memory, and gradually integrated into the longer-term memory as the text-processing proceeds. Hence, it may be reasonable to assume that when a reader encounters a pronoun, s/he will create a new information gathering point (a new 'node') somewhere in a temporary memory under which the information related to the pronoun is gathered, and from which the information is gradually integrated into longer-term memory.

Although this thesis does not deal with issues related to a writer's perspective, it is worth posing a question here: for those deferred completion cases, does the writer really assume that the referent of the pronoun is unambiguously identifiable when a reader encounters the pronoun ? It is rather hard to imagine that the writer assumes that the referent of the pronoun is salient enough for a reader to identify it uniquely at the encounter point associated with the pronoun. It is more likely the case that the writer assumes not the encounter point of the pronoun but a somewhat later point as the completion point. Such deferred completion cases appear to be another kind of 'wait-a-second' use of a pronoun, a phenomenon I proposed in 6-4.

7-4. Suspension of interpretation caused by structural dependency

It is now necessary to take account of the suspension of interpretation caused by structural dependency in accounting for, in particular, initial-ADV and 1PSP-initial-DS cataphora. Structural dependency (modification), such as an initial adverbial being dependent on the main clause, involves suspension of the interpretation of the dependent (modifying) element until the modified independent element is processed and interpreted, as Sørensen noted:

> ... if the subordinate clause comes first, we are forced to be attentive, we are kept in suspense till we have heard or read the main clause; this suspense reflects a special form of cohesion which makes it possible to defer the question of identity (of coreferentiality or non-coreferentiality) till the main clause.
> (Sørensen, 1982:146)

Ariel also talks about "the interpretative dependency of the pronoun on the full NP" (Ariel, 1990:160):

> ... first-mention (backwards anaphora) pronouns, which are indeed dependent on the matrix for their interpretation (unlike other cases of backwards anaphora) also require a high (syntactic) dependency of the pronoun clause on the antecedent clause.
> (Ariel, 1990:161)

The suspension of interpretation is considered to involve the holding in temporary short-term memory of information obtained through processing the initial adverbial, as Stockwell suggests:

> In the reading process, cataphoric markers will simply be seen as incomplete referring expressions, and will be held in short-term memory until a referent is found for them.
> (Stockwell, 1995:168)

Different types of initial adverbials may involve different degree of dependency. Givón noted that "participal ADV-clauses are more integrated, functionally and structurally, into their main clause" and they "tend to have the same-subject as their main clause" (Givón, 1993:838). For instance:

[A011 57] Predicting >*REF=102 he* will fare "much better than anyone expects" in Sunday's Maine caucuses, *(102 California Gov. Edmund G. Brown Jr. 102)* Tuesday compared his Democratic presidential rivals to "two peas in a pod."

The semantic subject (agent) of "Predicting" in the initial participal clause is the same as that of the main subject, 'California Gov. Edmund G. Brown Jr.'. In this respect, participal ADV-clauses are structurally and interpretatively integrated into the main clause.

Some other type of initial adverbials, on the other hand, are considered to be more loosely integrated into the main clauses, akin to co-ordination. For instance,

[A015 54] When *Cleveland firemen* arrived at the house, *they* found Robert lying behind the bed in Mrs. Dumas' upstairs bedroom.

The initial adverbial clause here contains a full NP (highlight trigger) "Cleveland firemen", the referent of which is mentioned by the 3PP (highlight-maintainer) *they* in the main clause subject. In this case, the main clause contains the maintainer of the highlight which is signalled to trigger by the full NP "Cleveland firemen" in the initial adverbial. In this respect, the initial adverbial clause is semantically (referentially) more independent than the initial participal ADV-clause in [A011 57]. The anaphoric relation between "Cleveland firemen" in the initial adverbial clause and 3PP *they* in the main clause indicates loose dependency of the initial adverbial clause on the main

clause. Yet no matter how loose the dependency is, the initial adverbial marker (in this case "When") at the beginning of the adverbial indicates that the adverbial cannot be independent on its own; it has to be complemented by the main clause. The main clause, on the other hand, does not need to be complemented, at least formally, by other clauses.

In any case, because of the suspension of interpretation caused by the structural dependency, the critical (or final) interpretation of the initial adverbial is suspended until the main clause is processed and interpreted. Hence the interpretation of a pronoun within the initial adverbial is suspended likewise.[62]

Since a reader tends to perceive a full NP in the subject position as a highlight-trigger (cf. Centering theory's claim in 2-8), the pronoun in the initial element (ADV or DS), which is dependent on the following highlight-trigger (main subject) may not be perceived as a signal to maintain the highlight on a previously focused entity. In this case, a reader does not perceive the continuation of the highlight on the previously focused entity (highlight-continuity). It can be said that when a reader becomes quite sure about the referent of the pronoun in the initial element (completion point)[63], the more a reader perceives dependency of a pronoun onto the following main subject full NP, which signals to trigger a highlight, the less s/he will perceive highlight-continuity. The reader, in this case, will perceive the highlight (focus) triggered by the main subject full NP and judge the pronoun reference as **cataphoric**, presumably as in the 3PP in [A034 102] below.

---

[62] It may be the case that the initial-ADV cataphora pattern has developed based upon the structural suspension caused by the structural dependency of an initial adverbial on the main clause.

[63] It can be assumed that in a preferred construction of cataphora, a reader tends to become quite sure about the referent of the initial pronoun usually towards the end of the sentence, after the main subject full NP is encountered.

[A034 101] In some countries, the best thing is to be a ski racer," **(119 Kiesel 119)** said Friday.

[A034 102] Although *>REF=119 he* didn't mention any countries by name, *(119 Kiesel 119)* clearly was referring to the Russian team which ...

Even if a reader perceives the dependency, if the coreferential main subject is a pronoun, which signals highlight-maintenance, a reader will perceive **highlight-continuity**. Hence the pronoun reference is judged as **anaphoric**, as in the 3PP *she* in the initial adverbial (marked as anaphoric <REF=) below:

[A044 35] And whether *<REF=30 she* liked them or not , *<REF=30 she* rarely hesitated to say exactly how <REF=30 she felt about them.

7-5. Immediate initiation and rapid completion of interpretation

Although the critical / final interpretation of initial adverbials is suspended until the main clause is processed, a reader does not simply suspend the interpretation of the pronoun in the initial adverbial until the main clause is processed but tries to predict the referent as soon as possible, even if s/he is unsure about success. Wales notes:

> The crucial question for cataphoric reference ... is how long a reader ... must 'wait' for a supposedly coreferential ... NP. ... We as addressees are looking for more information, in effect, about this person in the universe of discourse, in particular their identity, not waiting for a master NP to appear. And this information gradually is supplied to us, from our interpretation of many co-textual clues and inferences.
> (Wales, 1996:40)

In other words, as soon as a reader encounters a pronoun, s/he initiates the interpretation (**immediate initiation of interpretation**), and tries to complete it as quickly as possible (**rapid completion of interpretation**). The **immediate initiation** and **rapid completion** of interpretation is put forward by Sanford and Garrod (1989) as the **immediacy hypothesis** (originally proposed by Just & Carpenter, 1980):

> A strong version of this (*immediacy hypothesis) is the claim by Just and Carpenter (1990) that, eye fixations are maintained until the item has been processed to the deepest possible level. A weaker and equally interesting view, would be that completion takes place rapidly, even when there are arguable opportunities for deferral. This version underlies accounts of processing that assume temporary assignment with the possibility of later revision.
> (Sanford and Garrod, 1989:239)

With regard to the initial adverbial construction, as soon as a reader encounters a 3PP in the initial adverbial, s/he starts to make a prediction about the

referent, accessing the possible referent candidates, among which the most preferred

candidate is the one in focus (highlight), as Sanford and Garrod note:

> When a pronoun is encountered, it immediately triggers a process which is
> effectively seeking a match in terms of number and gender, where major
> candidate is that which is in focus.
> (Sanford and Garrod, 1989:254)

For instance, as soon as the 3PP he in [A022 17] is encountered:

[A022 16] Over the years, **DeLury** gained a reputation on both sides of the bargaining
         table as an effective labor leader and an innovative negotiator.
[A022 17] When **he ...**

a reader is not simply waiting for the further input but will make a tentative prediction

of the referent as 'DeLury', who is a number-gender-matched entity and also

highlighted.

[A022 16] Over the years, **(1 DeLury 1)** gained a reputation on both sides of the
         bargaining table as an effective labor leader and an innovative negotiator.
[A022 17] When **<REF=1 he** retired, **(1 DeLury 1)** was succeeded by ...

In this case, even though the following main subject is a coreferential full NP, the AT

analyst judged the 3PP as anaphoric, presumably because the referent could be

predictable with a high degree of certainty.

7-6. Summary and Discussion

After examining the borderline cases in terms of reference-direction judgement (anaphora/ cataphora/ indeterminate) in 7-2, we have observed that:

(i) The majority of the indeterminate cases (7 out of 10) exhibit the typical cataphora patterns (initial-ADV or 1PSP-initial-DS).

(ii) Borderline cases can be observed in the initial-ADV and 1PSP-initial-DS constructions not only in indeterminate examples but also in examples marked as either anaphoric or cataphoric.

Hence, so far as the AT data is concerned, it can be said that borderline cases typically involve the initial-ADV or 1PSP-initial-DS constructions, in which the coreferential full NP occurs as the main subject (including the subject of reporting clause). In the following discussion on the reference-direction judgement, I will mainly focus on the initial-ADV and 1PSP-initial-DS constructions.

In general, it can be said that if a reader perceives continuity of the highlight placed on the previously focused item (**highlight continuity**), s/he will judge the pronoun reference as anaphoric. The more a reader perceives highlight-continuity, the more s/he tends to judge the pronoun as **anaphoric.**

With regard to the initial-ADV or 1PSP-initial-DS constructions, there are two main factors which influence the highlight-continuity:

(i) The degree of certainty with which the referent of a pronoun is predictable before the coreferential main subject is encountered.

Due to the requirement of **immediate initiation** and **rapid completion,** a

reader usually makes a prediction of the referent of a pronoun in the initial element (ADV or DS) before s/he encounters the coreferential main subject (including the subject of the reporting clause). The more the referent of a pronoun is predictable for a reader before encountering the coreferential main subject, in other words the more a reader perceives highlight-continuity, the more s/he is likely to judges the pronoun reference as **anaphoric**. The less the referent of a pronoun is predictable for a reader before encountering the coreferential main subject, the less s/he perceives highlight-continuity, and the more s/he perceives the highlight triggered by the main subject full NP when s/he becomes certain about the referent of the pronoun (usually after encounter the main subject full NP); hence the more s/he is likely to judges the pronoun reference as **cataphoric**.

(ii) The extent to which a reader perceives the dependency of the pronoun in the
    initial element (ADV or DS) on the subject full NP in the main clause.

Due to the **structural dependency** involved in an initial adverbial or an initial DS, the interpretation of the initial element, at least the critical interpretation of it, is **suspended** until the main clause is processed. The more a reader perceives dependency of the pronoun in the initial element (ADV or DS) on the following main subject full NP, which triggers a highlight, the less s/he perceives highlight-continuity of the previously focused entity when s/he completes to resolve the pronoun; hence the more s/he tends to judge the pronoun reference as **cataphoric**. Even if a reader perceives the dependency, if the coreferential main subject is a pronoun, which signals highlight-maintenance, a reader will perceive **highlight-continuity**, and will judge the pronoun reference as **anaphoric**.

Figure 7.6.1 summarises how a 3PP in an initial adverbial or 1PSP in an initial

DS is judged in terms of the reference-direction in the AT.

Figure 7.6.1[64]

Main subject (MS) is coreferential (coref) with the pronoun in the initial adverbial or initial DS (Initial pronoun)		
(I) MS is a pronoun (highlight-maintainer)	MS is a full NP (highlight-trigger)	
	(IV) Previous highlight is coref with Initial Pronoun (**Borderline cases**)	(III) Previous highlight is **ambiguous** or not coref with Initial Pronoun (**Overriding Type**)	(II) Initial Pronoun is **First-mention**
Initial Pronoun is judged as **anaphoric**	Initial Pronoun is judged as **Indeterminate**	Initial Pronoun is judged as **cataphoric**

As shown above, the initial-ADV and 1PSP-initial-DS examples can be roughly

grouped into four types in terms of reference-direction judgement:

(I) Main subject is a pronoun (highlight-maintainer)

In the initial-ADV pattern, when the coreferential main subject is a pronoun,

that signals highlight-maintenance, a reader will perceive **highlight-continuity** from

the preceding text, and judges the pronoun reference as **anaphoric**. Yet occasionally

in the 1PSP-initial-DS pattern, a reader judges the pronoun reference not as

---

[64] Abbreviations in the Figure :
    'MS' : main subject. i.e. subject of the main clause of the current sentence.
    'coref' : coreferential.
'Initial pronoun': a 3PP in the initial adverbial or 1PSP in the initial DS.
'Previous highlight' : previously highlighted entity.

anaphoric but as **indeterminate**, as shown in (7.4) below, presumably due to the ambivalence of the function of the 3PP *he* in regular coreferential relation (antecedent) and in the highlight signal relation (anaphor), as discussed as discussed with (7.4) and (7.5) in 7-2-1.

(7.4)   *?<>REF=29 I* in [A030 29]

[A030 28] But without the Olympics, **(29 Leonard 29)** says **<REF=29** he probably
        would have gone to college and chosen a more conventional occupation.
[A030 29] **"?<>REF=29 I** wouldn't be a fighter today without the Olympics,"
        <REF=29 he said.

(II)   Main subject is a full NP (highlight-trigger) and Initial Pronoun is a **First-mention** case

For a First-mention pronoun, simply because it is impossible to talk about the continuation of previously highlighted entity, and the reader perceives the highlight triggered by the main subject; hence the reference must be judged as **cataphoric**.

(III) Previous highlight is **ambiguous** or not coreferential with the Initial Pronoun
        (Overriding Type)

When the previously highlighted entity is ambiguous (i.e. it cannot be uniquely identifiable) or is not coreferential with the initial pronoun, a reader cannot perceive highlight-continuity but s/he perceives the highlight triggered by the main subject; hence the reference is judged as **cataphoric**.

(IV)   Main subject is a full NP (highlight-trigger) and the Previous highlight is coreferential with the Initial Pronoun (**Borderline cases**)

As looked at in 7-2, when both anaphoric and cataphoric readings have certain evidence to support them in the borderline cases, there is a room for variability of reference-direction judgement according to the individual reader. The judgement can vary according to such factors as:

(i) The degree of certainty with which the referent of a pronoun is predictable before the coreferential main subject is encountered.

(ii) The extent to which a reader perceives the dependency of the pronoun in the initial element on the subject full NP in the main clause.

(iii) The extent to which the preferred constructions of cataphora are internalised in each reader's interpretative strategy.

In summary, the AT data reveals that in certain conditions, the reference-direction judgement of a pronoun cannot be made by a clear-cut binary fashion but involves a triple-choice (anaphoric/cataphoric/indeterminate), or it could involve a cline between the anaphoric pole and the cataphoric pole.

Discussion of the view on the notion of reference-direction (anaphora/cataphora) addressed by Stockwell (1995) and Wales (1996)

As noted in 7-1, some cognitive discourse theorists claim that the notion of reference-direction (anaphora/cataphora) can be dispensed with altogether from the point of view of on-line text processing. Stockwell (1995) notes:

> Incidentally, an on-line treatment of reference processing clearly precludes the notion of cataphora (forward co-reference). This is again a structuralist concept derived by treating the text as an object which can be worked through to and fro by the linguist. In the reading process, cataphoric markers will simply be seen as incomplete referring expressions, and will be held in short-term memory until a referent is found for them.
> (Stockwell, 1995:168)

I believe that Stockwell is right in saying that the mental representations of cataphoric markers "will be held in short-term memory until a referent is found for them".
I also have suggested several times that a cataphoric pronoun needs to be assumed to create a temporary information gathering point (a temporary node) in short term memory.

Yet I would argue that contrary to his claim, cataphoric pronoun references cannot "simply be seen as incomplete referring expressions" such as deferral completion cases, because of the following reasons: Firstly, cataphora involves typical coreferential patterns (initial-ADV and IPSP-initial-DS), which enable a reader to have an anticipation that the coreferential NP is likely to occur in the following main subject. Whereas in a deferral completion case, a reader does not have such anticipation. Secondly, although a reader may not perceive a highlight-continuity in both genuine cataphora and a deferral completion case, s/he may perceive the

highlight triggered by the following main subject in cataphora, whereas s/he does not perceive such highlight in deferral completion cases.

We can define anaphora to be "a signal to continue the existing attention focus already established" (Cornish, 1996:22). In other words, when a pronoun is anaphoric, a reader tends to perceive the continuity of highlight (focus) placed on the previously highlighted (focused) entity. The borderline cases in terms of reference-direction (anaphora/cataphora) reveal that in certain conditions in which the pronoun in the initial (preposed) element co-refers with the following main subject, the reference-direction (anaphora/cataphora) judgement tends to involve a triple-choice (anaphoric / cataphoric /indeterminate), or it could involve a cline between the anaphoric pole and the cataphoric pole. In such conditions, we can say that the more a reader perceives the continuity of highlight on the previously focused entity, the more s/he is likely to judge the pronoun reference to be anaphoric; and the less the reader perceives the highlight-continuity, the more s/he perceives the highlight triggered by the main subject when s/he becomes certain about the referent of the pronoun (completion point), hence the more the reader judges the pronoun reference to be cataphoric. In this respect, the reference-direction judgement (anaphora/cataphora) of a pronoun tends to reflect the way in which a reader perceives highlight (focus) at the completion point of the pronoun resolution, hence it has a certain cognitive rationale. Therefore, contrary to Wales's claim (1996:27), the notion of reference-direction cannot be to dispensed with "virtually altogether" in on-line text analyses.

Chapter 8: Conclusion

8-1. Summary of the thesis

The analysis of the Anaphoric Treebank and other corpus data reveals certain regular coreferential patterns of cataphora. The regularity is by no means rigid as syntactic constraints, but it is significant enough to show that these preferred constructions of cataphora can be counted as an interpretative schema acquired by a native speaker of English, as Van Hoek notes:

> Backwards anaphora must rather be specifically sanctioned by a constructional schema or network of interrelated schemas, which are acquired by speakers through exposure to actually occurring expressions.
> (Van Hoek, 1997:111)

On the basis of the corpus data, this thesis examines various claims on cataphora made by generative approaches and cognitive discourse theories (Mental representation theories).

Generative approaches

(1) Problem of acceptability judgement

I carried out informant tests to examine Reinhart's claim for the syntactic distinction between initial verb-phrasal adverbials and initial sentential adverbials. As a result, it became clear that the acceptability judgements made by Reinhart considerably differs from the judgements made by 15 native speakers.

I would claim that this result indicates the inadequacy of a narrow restriction of data to invented examples in generative approaches, which is a consequence of excluding empirical phenomena from the object of inquiry due to the influence of Platonic idealism.

(2) Also the lack of understanding of the scope of application of theories can be observed in Reinhart as well as in Binding theories.

Reinhart's C-command model and Binding theory are almost incapable of accounting for the instances of cataphora observed in corpus data.

(3) Through examining Carden's 'Block Forward cases', it became clear that it is inadequate to confine oneself to the sentential level approach for studying pronoun references.

Cognitive discourse theories (Mental representation theories)

It became clear that First-mention and Overriding type cataphora cannot be accounted for by the hierarchical models such as Ariel's accessibility scale or Gundel et al's Givenness hierarchy (6-1, 6-2).

The empirical data (First mention cataphora and Overriding-type) clearly

shows that there is such a phenomenon that can be called cataphora, contrary to those

sceptical views on cataphora among cognitive discourse analysts (Kuno, Cornish) (6-

3).

These empirical data indicate that the use of the pronoun is not restricted to

anaphoric use. In other words, a pronoun reference can be made not only for already-

mentioned discourse entity or an already-focused (salient) discourse entity but also for

a brand-new (never-heard-before) entity.

The borderline cases in terms of reference-direction (anaphora/cataphora)

reveal that in certain conditions in which the pronoun in the initial (preposed) element

co-refers with the following main subject, the reference-direction

(anaphora/cataphora) judgement tends to involve a triple-choice (anaphoric /

cataphoric /indeterminate), or even a cline between the anaphoric pole and the

cataphoric pole.

Through analysing the borderline cases in terms of (i) Suspension of

interpretation caused by structural dependency and (ii) Immediate initiation and rapid

completion of interpretation of pronouns, I argue that the reference-direction

judgement of a pronoun (anaphora/cataphora) tends to reflect the way in which a

reader perceives focus of attention (highlight) when s/he completes to resolve the

pronoun, hence it has a certain cognitive rationale. It is not, therefore, reasonable to

dispense with the notion of reference-direction "virtually altogether" in the on-line

text analysis. I also argue that, contrary to the claim made by Stockwell(1995:165),

cataphoric pronoun references cannot "simply be seen as incomplete referring expressions" such as deferral completion cases.

8-2. Further discussions

From an addresser(writer/speaker)'s perspective, I suggest that it may be necessary to assume what might be called 'wait-a-second' use of pronoun to account for the cataphora examples as well as deferring completion examples.

From a reader's perspective, in discussing genuine cataphora and deferring completion cases, I propose to assume that a pronoun creates a temporary information gathering point (a new node) somewhere in reader's short term memory. This point is also made by some other linguists such as van Deemter (1989) or Stockwell (1995).

> ... a kataphoric (sic) element introduces an incomplete discourse entity, to be completed by subsequent material under certain conditions.
> (van Deemter, 1989:113).

> In the reading process, cataphoric markers ... will be held in short-term memory until a referent is found for them.
> (Stockwell, 1995:165)

It seems to be the case that this function, the creation of a temporary node in short term memory, provides the basis not only for genuine cataphora and deferring completion cases but also for ordinary anaphora (highlight-maintenance signal). The difference between ordinary anaphora and cataphora or deferring completion

cases, in this respect, is the matter of degree of gap between initiation point and completion point of interpretation.

This leads to a further question: the creation of a temporary node in short term memory takes place not only in the use of pronouns but also in the use of proper names and lexical NPs?

So far as ambiguous cases are concerned, I would speculate so. As Sanford and Garrod note:

> It is a truism that all processes take time, and so it is necessary to discriminate between the onset of processes and the point of their completion
> (Sanford and Garrod, 1989:239)

During the gap between the initiation point and the completion point of interpreting an NP, the information obtained through processing the co-text of the NP needs to be held in some kind of temporary location of memory until the interpretation is stabilised.

Hence, it may be reasonable to assume that when a full NP cannot be resolved immediately, the reader will create a new information gathering point (a new 'node') somewhere in a temporary memory under which the information related to the full NP is gathered, and from which the information is gradually integrated into longer-term memory.

In this respect, the process of interpretation can be considered to be the process of making linkage between the newly created temporary node and the existing network of longer-term memory; or it can be the process of integrating the newly

created temporary node into the existing network of nodes in the longer-term

memory.

8-3. Limitation of thesis and Remained tasks for future studies

[1] This thesis has focused on cataphoric pronoun references from a receiver's point
of view, and to a less degree discussed it from an addresser's point of view.

[2] This thesis has not dealt with the socio-pragmatic aspects or psycholinguistic
aspect of pronoun references.

[3] This thesis is largely concerned with singular human personal pronouns. Other
types of personal pronouns (non-human or plural pronouns) are less focused in
this thesis.

[4] This thesis has not analysed cataphoric pronouns according to
subjective/objective/possessive/reflexive form.

[5] This thesis has mostly treated initial adverbials as a single category. This is
because most of the theories tested in this thesis concern syntactic subordination
or dependency, and hardly any of them analyse initial adverbials analytically.
Ideally 'Fine-grained' functional analyses of initial adverbials should be done
with discourse data.

[6] Mainly because of the limitation of time, this thesis could not carry out the
following research tasks:

* Testing Centering claim (see 2-8) on the AT data.

* To discuss Cognitive grammar approaches presented by Langacker (1987, 1991)
and Van Hoek (1997).

* To discuss cataphora and Discourse representation theory (DRT)

  Some DRT researchers like van Deemter (1989) try to develop algorithms to

account for cataphora in the view that cataphora "introduces an incomplete discourse entity". This thesis could not deal with van Deemter's proposal, but I believe that his approach is in the right direction.

* Careful psychological (replacement) tests for Overriding/ambiguous type of cataphora are needed.

[7] This thesis deals with only news reportage texts. Cataphora in other genres of text remain to be investigated.

[8] As discussed in 6-2, we cannot simply satisfy ourselves by finding examples that cannot be accounted for by the Givenness hierarchy (Gundel et al, 1993) or Accessibility scale (Ariel, 1990). We need to pay attention to how much effort a processor (reader) needs in order to retrieve the referent of a pronoun. The study of the retrieval cost for cataphoric use of pronouns is a task for future research.

8-4. Final remarks

This thesis is an attempt to demonstrate the usefulness of a corpus with discourse annotation such as the AT in the study of pronoun reference. The discourse annotation of the Anaphoric Treebank (AT), where the analytic judgements of native speakers are recorded through the annotations, could reveal that in certain conditions in which the pronoun in the initial (preposed) element co-refers with the following main subject, the reference-direction (anaphora/cataphora) judgement tends to involve a triple-choice (anaphoric / cataphoric /indeterminate), or it can involve a cline between the anaphoric pole and the cataphoric pole. The AT cataphora data also provide psychological evidence that a cataphoric reading can be made not only for the genuine cataphora cases (First-mention and Overriding type) but also for other Already-mentioned cases.

Finally I must stress that although it is still in refinement stage, the Anaphoric Treebank contains extremely rich information that can give considerable benefits to the empirical research of discourse anaphora in general. The present work is just one small outcome of this fruitful research resource.

Appendix 4-1-1. Backwards anaphora samples cited in Carden (1982)

(4) Eddie Mathews sacrificed Mantilla to second, and Hank Aaron was walked intentionally. That brought up Joe Adcock. On Haddix second pitch, Adcock swung and sent a towering drive that disappeared over the right-field wall. Haddix had lost his perfect game, his no-hitter, and the ball game.

 ...

It should have been a three-run home run, but in *his* joy, *Aaron* cut across the field after touching second base.
--- (Source is unknown)

(5) Jonathan Spence, professor of history and recently appointed replacement for Giamatti as director of the humanities division said, 'I'd rather have someone who's done some extensive university teaching and research than someone who hasn't. He added that 'a good president will have advisors who must help him.'

 Administrator-Scholar

While *he* hadn't read the Gifford article, *Associate Dean of Yale College Martin Griffin* said that 'the best administrators are scholars', and that. . .
 (*Yale Daily News*, 31 Jan. 78)

(6) 'Are we all here?' said he [Gandalf], handing *his* sword back to *Thorin* with a bow.
 (J. R. R. Tolkien, *The Hobbit*: 77)

(7) Mrs Tabitha became more and more distracted, and mewed dreadfully. While *their* mother was searching the house, *Moppet and Mittens* had got into mischief.
 (Beatrix Potter, *The Roly-Poly Pudding*: 10)

(8) To the Editor:
 After *his* recent election as Republican national chairman, *Bill Brock* said...
 (Letter, *New York Times (NYT)*, 28 Jan. 77)

(9) [There is also]·the fact that, as far as you can trust *him*, *Sturtevant* says that. . .
 (*LS*, g June 79)

(10) .While an armed host lies before our doors, we look on you as foes and thieves. It is in my mind to ask what share of *their* inheritance you would have paid to *our kindred*, had you found the hoard unguarded and us slain.
 (J. R. R. Tolkien, *The Hobbit*: 270)

(11) Details of *Her* Death Fill the Day For Family of *the Latest Victim*
 (headline, *NYT*, 3 Aug. 77)

(12) When *she* was five years old, *a child of my acquaintance* announced a theory that she was inhabited by rabbits.
 (*NYT*, 6 Nov. 78)

(13) Because *they* wanted to know more about the ocean's current, ***students in the Science Club at Mark Twain Junior High School of Coney Island*** gave ten bottles with return address cards inside to crewmen of one of New York City's sludge barges.

                    (*NYT*, 26 June 76)

(14) We are . . . lawyers who go into court to . . . return to *her* classrooms ***a pregnant girl*** illegally suspended from school . . .

          (advertisement, Children's Defense Fund, Nov. 78)

(15) Now, while the flame they watch not towers

        Above the soil they trod,

        Lads, we'll remember *friends of ours*

        Who shared the work with God.

          (Housman, '1887', *Shropshire Lad*)

(17) . . . there are six legally operated and licensed poker cardrooms . . . As *its* major source of income, ***each club*** collects a playing fee from the players every half hour. . .

         (*Social Problems*, 28: 557 (1977))

(18) Al Bowling and Tom Rentschler. . . Bowling. . . Rentschler ... In *his* own way, however, ***each man*** is petitioning for the same kind of Administration.

         (*NYT*, 21 Jan. 77)

(19) Scrimshawing took time. And, once *his* ship had reached the whaling grounds, time was something ***every whaleman*** had a great plenty of.

         (Shapiro, *The Story of Yankee Whaling*: 291 (1959))

(20) When *their* government tenure ends, ***many officials*** simply move to new offices.

         (*NYT*, 23 Dec. 77)

(21) Did you know that when *their* wives leave *them*, ***two men in five*** go bananas?

            Lanniglan's Rabbi. TV, fall 1977

            (collected by T. Dieterich)

(22) No matter how innocent *he* may be in his inner soul and in his motivations, ***the effective mathematician*** is likely to be a powerful factor in changing the face of society.

         (Weiner, *Ex-Prodigy*: 189-190)

(23) Unless *he*'s very hungry and therefore desperate, chances are ***a shark*** will not attack unless . . .

         (*Yachting*, Dec. 75)

(24) The Washington Area Women's Centre:

A Case Study in Organizational Growth and Change:
Despite *their* egalitarian ideologies and objectives, ***voluntary organizations and social movement organizations (s.m.o.s)*** have not been immune to bureaucratization and organizational transformation.
(paper submitted for refereeing, June 76; author's name deleted)

Appendix 4-1-2. Cataphora data presented by Kanzaki (1994)

* Cataphoric pronouns and supposed antecedent NPs (according to Kanzaki's judgement) are typed in bold and italics.
* For further source of refernce, see Kanzaki (1994).

Antecedent is indefinite NP

(45) Asked if *he* was "certain" or just "highly confident" that Bush had been targeted by Saddam Hussein, *a senior U.S. intelligence official* tersely replied, "We're certain. Al-Ghazali was tasked specifically to kill President Bush."
        Time 5 July, 93'   Kanzaki (1994:101)

Previously mentioned antecedent but overriding competing candidate case

(48) 'Sit down,' he said, gesturing at a chair. 'Sit down.' Benjamin sat. 'Well,' Mr. Robinson said, raising his glass. 'Here's to you and your date.' As *he* was drinking *Benjamin* looked over the rim of his glass at Mrs. Robinson.
        C. Webb, *The Graduate*, 95 Kanzaki(1994:103)

Examples appearing at the beginning of the text or paragraph. and cannot invite anaphoric reading.
        Kanzaki (1994:104-107)

(49) When *he* was a student at U-M in the late 1960s, *John White* brought his carving tools to Pretzel Bell Restaurant and carved "Fiz," his nickname, into the tabletops.
        State News, 18 Apr. '85

(50) Back when *she* was a student at Harvard. *Masako Owada* introduced some of her friends to a card game called Emperor.   Time, 7 Jun. '93

(51) Before *he* left for the Tokyo summit of world leaders this week, *President Clinton* had to decide whether to bomb Iraq in retaliation for an attempted assassination of former president George Bush;...
                    Newsweek, 12 Jul. '93

(52) Now that *he* is leaving Los Angeles as the dominant athlete of the Olympic Games after winning four gold medals, *Carl Lewis* wants a chance to continue life as plain old Carl Lewis.   State News, 13 Aug. '84

(53) While *he* lived, few Japanese had heard of *Atuhito Nakata*, but for his country he was an important symbol.   *Newsweek*, 10 May '93

(54) Whatever *her* failures or *her* faults, **Margaret Thatcher** has been Britain's outstanding peacetime leader of the twentieth century.
The Economist, 29 Apr. '89

(55) For 56 days, while *his* rebel army fought the Angolan government for the city of Huambo, **Jonas Savirnbi** kept his plans to himself.
Newsweek, 22 Mar. '93

(56) For half *her* life, **Paloma Herrera**, 17, has been on her toes. The youngest member of America Ballet Theater's corps, the Buenos Aires beauty is an ascending star, . . . *Time*, 31 May '93

(57) Since *his* election, **Bill Clinton** has more than demonstrated his brains, energy and rapid grasp of complex issues. *Time*, 25 Jan. '93

(58) In *his* Inaugural Address, **Bill Clinton** described Washington as "a place of intrigue and calculation." *Time*, 1 Feb.

(59) Although *he* had rejected Jespersen's attribution of language change to 'desires' or 'needs' in his review of Jespersen's *Language*, **Bloomfield** retained his belief in Jespersen's thesis that historical change is progressive.'
[Falk 1992: 478]

(60) In *his* study of "small talk" **Schank** (1977) formalized some of the possible relationships that utterances may have with one another.
[Reichman 1978: 287]

(61) In *his* discussion of obviation in Menomini, **Bloomfield** formulates this as follows: ...
[Dik 1989: 274]

(62) In *his* analysis of an example of written text, **van Dijk** proposes that its topic can be represented as a proposition that is nontrivially and jointly entailed by the ordered sequence of propositions expressed by the sequence of sentences in a text.
[Blakemore 1988: 234]

(63) In 1756, the year after the publication of *his* Dictionary, **Dr Johnson** was in difficulties, ...
[Weekley 1952: 21]

(64) In Chapter 1 of *his* forthcoming book, Deep and Surface Grammar (Cambridge, Mass.: M.I.T.Press), **George Lakoff** defines more precisely the notion of precyclic rule ... [Ross 1969: 190]

(65) During much of *his* professional life, **the Danish linguist Otto Jespersen** (1860-1943) was well known and well received in the United States.
[Falk 1992: 465]

(66) Though *he* tried not to think of it, at twenty-nine ***Tommy Castelli's*** life was a screaming bore.
B. Malamud, *The Prison*, 97

(67) When it came to concealing *his* troubles, ***Tommy Wilhelm*** was not less capable than the next fellow.
S. Bellow, *Seize the Day*, 3

(68) It was Sunday, and according to *his* custom on that day, ***McTeague*** took his dinner at two in the afternoon at the car conductor's coffee joint on Polk Street.
F. Norris, *McTeague*, 1

(69) For twenty years, every autumn since *her* marriage, ***Margaret Fleming*** had watched the leaves from this window; and always it had seemed to her that they were a part of her life which she held precious.
E. Glasgow, *'The Difference*," 965

(70) ***Her*** doctor had told ***Julian's mother*** that she must lose twenty pounds on account of her blood pressure,
F.O'Connor, *Everything That Rises Must Converge*, 3

Appendix 4-2-1: First-Mention Cataphora in the Anaphoric Treebank.

Each example is presented with the sequential number of the example (enclosed by angle brackets), source reference (enclosed by square brackets), type of cataphora (e.g. initial-ADV / 1PSP-initial-DS / and Other types), followed by the example.

<1>   [A004 89]   Other type (unequal conjoined structure)
"A lot of players dread thinking about *it*, but *that day always comes in baseball when you have to go out and get a real job*," he said.

<2>   [A005 53]   1PSP-initial-DS
"*I* think this contract is a major step forward," said *Paul Silas*, president of the Players Association.

<3>   [A005 60]   1PSP-initial-DS
"*I* have to give a lot of credit to my good friend Ben Crenshaw," *George Burns* said after scoring his first official individual victory Sunday in the Bing Crosby National Pro-Am.

<4>   [A005 114]   initial-ADV
Shortly after *he* became the FCC's chairman, *Charles D. Ferris* told Congress the licensing procedure is an "anachronism."

<5>   [A006 35]   Other type (unequal conjoined structure)
I can not promise *it*, but we will try for *$1,000*."

<6>   [A009 66]   1PSP-initial-DS
"Biggest thrill of *my* life," said *John Montgomery*, a 5-foot-6, 160-pound transfer student who blasted home the winning goal, a 45-footer at 5:09 of the overtime.

<7>   [A010 105]   1PSP-initial-DS
"Am *I* shocked ?" *House Speaker Thomas P. O'Neill* asked rhetorically.

<8>   [A011 57]   initial-ADV
Predicting *he* will fare "much better than anyone expects" in Sunday's Maine caucuses, *California Gov. Edmund G. Brown Jr*. Tuesday compared his Democratic presidential rivals to "two peas in a pod."

<9>   [A012 41]   initial-ADV
Several individual members who were questioned refused to answer policy questions and when asked what *he* hoped to achieve from the visit, *Greg Dill, of Evanston, Ill.*, said "to listen, to hear and to get insights."

<10> [A014 53]   1PSP-initial-DS
"*I* 'm just praying for a little white stuff," said *Joe Santa Fe*, owner of Chrysler Viking Snow Plowing Co. here.

<11> [A014 93]   Other type (1PPP-initial-DS)
"*We* only transported one person to the hospital by ambulance for treatment because of the gas," said *fire department* spokesman Dale Everett.

<12> [A015 42]   1PSP-initial-DS
"*I* heard him yell, Where's grandma, where's grandma ?" ' said *his mother*, Jacqueline Hodge, 25.

<13> [A017 9]   1PSP-initial-DS
"It gives *me* a great deal of pleasure to bring Roger back to the organization," said Coach *Bobby Leonard*.

<14> [A017 42]   Other type
"At *its* broadest point, I 'd say *the water* is about 1,000 feet wide.

<15> [A018 24]   1PSP-initial-DS
"*I* figured there 'd be 200 at the most," said *South Windsor Republican town chairman Richard Rittenband*, who attributed the crowd of nearly 600 at a Lincoln Day fund-raiser Monday night to Bush's success in the Iowa caucuses and emergence as a front-runner in the polls for the Republican nomination.

<16> [A018 26]   1PSP-initial-DS
"*I* never figured George Bush would be second to anyone in Connecticut," said GOP *National Committeeman John Alsop*, a Bush supporter attending the East Windsor affair, stating that Bush's Iowa victory had improved his competitive position in the state.

<17> [A018 56]   1PSP-initial-DS
"Using the '60s as a base, *I* don't feel safer ; but I feel safer than in the '70s," is how *one guard* put it.

<18> [A019 11]   1PSP-initial-DS
"*I* don't understand, the students don't understand, no one understands," said *Richard Dillon*, principal of Stowe High School, which both Michael Huard, 19, and Lesley Phillips, 16, had attended.

<19> [A020 30]   1PSP-initial-DS
"The normal four racers would have been Peter Wirnsberger, Josef Walcher, Harti Weirather and *myself*," said *downhiller Werner Grissmann*.

<20> [A022 4]   1PSP-initial-DS
"*I* knew John DeLury," said *Mayor Edward I. Koch* on learning of his death.

<21> [A025 13]   1PSP-initial-DS
"*I* will go on fighting for the athletes of my country," said *Henry Hsu*, the Taiwan member of the International Olympic Committee (IOC).

<22> [A025 69] 1PSP-initial-DS
"*I* think Bailey's dunk woke up 20,000 people who were very nervous and thought they were going to lose," said *Atlanta Coach Hubie Brown* after the Hawks dropped a 93-86 National Basketball Association game Wednesday to the Sonics.

<23> [A027 52] Other type (1PPP-initial-DS)
"Maybe that is *our* secret," said Dick de Vroomen, *the team* leader.

<24> [A029 43] 1PSP-initial-DS
"*I* thought we had the game won in regulation," said *North Carolina Coach Dean Smith.*

<25> [A029 104] 1PSP-initial-DS
"*I* shook hands with Robert (Kennedy) one time in Boston," *one woman* told him.

<26> [A031 85] 1PSP-initial-DS
"The soldiers who came to *my* house were good Moslems," said *Abdul Rauf*, a radio repairman.

<27> [A033 32] initial-ADV
Amid demands for *his* resignation, *Democratic state Rep. Russell J. Reynolds* apologized Friday for using a racial slur in a response to a survey.

<28> [A033 97] initial-ADV
In *his* brilliant career, *East German luger Dettlef Guenther* has logged more than 10,000 runs on the frigid, writhing conduits built for the spartan sleds and riders who participate in one of the most elemental of winter sports.

<29> [A035 59] Other type
"Once last year, after *we* lugged all the equipment up, *three of us* just fell on the snow and looked down at the clouds below.

<30> [A037 67] 1PSP-initial-DS
"*I* was pleasantly surprised," *Manager Jim Frey* said at the end of a two-hour workout for 22 players who showed up for the start of a special rehabilitation and instructional camp at Terry Park.

<31> [A037 94] 1PSP-initial-DS
"*I* 've never seen Rocky," *Dave Thompson*, Clemente's manager, said at Monday's weigh-in.

<32> [A038 34] 1PSP-initial-DS
"Maybe it's *me*," *an embittered Coach Bobby Kromm* said.

<33> [A039 3] 1PSP-initial-DS
"*I* believe this may be the first case filed in the United States dealing with the constitutionality of police chaplains ' positions," says *Charlotte assistant attorney Richard D. Boner.*

<34>   [A042 11]   1PSP-initial-DS
"*I* would suppose business is looking over the valley," said *a Commerce Department analyst* who asked not to be identified.

<35>   [A042 75]   1PSP-initial-DS
"They knew each other well enough, *I* guess," *Weissmandl* said.

<36>   [A044 2]   Other type (1PPP-initial-DS)
"The cost of funds is climbing so rapidly *we* haven't any other choice," David Reed, senior vice president of *California Federal's loan division*, said Wednesday.

<37>   [A045 100]   1PSP-initial-DS
"*I* was pleased with our effort," said *State coach Jim Hatfield*, whose team fell to 12-13 overall and 6-11 in the SEC.

<38>   [A046 11]   initial-ADV
In a dissenting opinion for *himself* and Justices William J. Brennan Jr. and Thurgood Marshall, *Justice Harry A. Blackmun* objected to "direct aid for ostensibly secular purposes."

<39>   [A047 67]   Other type (Right dislocation)
They are shrouded in mystery, *these hockey playing visitors from the Soviet Union*, and there is irony in that.

<40>   [A047 91]   initial-ADV
For *his* part, *American Coach Herb Brooks* has his own opinions on the Soviet style and its practicality for North American players.

<41>   [A048 6]   1PSP-initial-DS
"*I* think you have seen the Big Ten champions tonight," *MSU Coach Jud Heathcote* said of the Hoosiers.

<42>   [A048 10]   1PSP-initial-DS
"*I* think the most important part of the game was when we were down 56-50 and we scored eight straight points and went ahead," *Coach Bobby Knight* said of the Hoosier comeback.

<43>   [A049 49]   Other type
A slump in fourth quarter camera sales and continuing marketing problems with *its* instant movie system has dropped *Polaroid Corp.*'s 1979 earnings to less than half last year's levels.

<44>   [A050 63]   1PSP-initial-DS
"It was a terrible tragedy, but *I* think no one will ever know what caused the fire," said *Dorothy Spicer*, a member of the jury that spent about 2.5 months wrestling with the case.

<45>   [A057 39]   Other type
"Not the easiest place to sign a treaty on, is *it* ?" he said, running his hand over *the rough surface of the table*.

<46>   [A057 54]   1PSP-initial-DS
"*I* really don't know," answered *the guide*.

<47>   [A059 58]   1PSP-initial-DS
"*I* was very pleased that we kept our poise tonight," said ***Kansas Coach Ted Owens***.

Appendix 4-2-2: Previously-mentioned cataphora in the Anaphoric Treebank.

Each example is presented with the sequential number of the example (enclosed by angle brackets), source reference (enclosed by square brackets), type of cataphora (e.g. initial-ADV / 1PSP-initial-DS / and Other types), followed by the example.

<u><1> *it* in [A003 19]</u>   initial-ADV

[A003 14] The vice president of **the West German Olympic Committee** said Sunday he believes the Summer Olympics can not be held in Moscow this year.
[A003 15] "To hold the Olympic Games in Moscow is no longer possible," Max Danz, 71, said in an interview with the Hamburg-based Welt Am Sontag newspaper.
[A003 16] Danz said he was against a boycott of the Games.
[A003 17] Instead, he suggested that the Games be delayed a year, but still be held in Moscow.
[A003 18] The official did not elaborate on why he thought it would be impossible to hold the 1980 Games in Moscow.
[A003 19] Although *it* has expressed support for holding the Games, *the German Olympic committee* has taken no final stand on President Carter's proposal to move, postpone or cancel the Games unless the Soviet Union removes its troops from Afghanistan by Feb. 20.

<u><2> *she* <3> *her* <4> *her* in [A003 27]</u>   initial-ADV (3)

[A003 23] Jacobson owned the Upper East Side apartment building where he, Tupper and Miss Cain lived and where he and Miss Cain ran a modeling agency.
[A003 24] "You better have two bodyguards," Tupper had said," and they better guard you 24 hours a day, and they better have big guns . "
[A003 25] The surprise tape recording Wednesday was but the latest twist in the long-running case.
[A003 26] There have been death threats against witnesses, missing witnesses and East Coast manhunts for neighbors who haven't been seen since the 1978 killing. ·
[A003 27] As *she* listened in court with headphones to *her* two former lovers fighting over *her, the slender, honey-blonde Miss Cain* started to cry.

<u><5> *I* <6> *I* <7> *my* in [A003 105]</u>   1PSP-initial-DS (3)

[A003 100] **Lloyd** had said she wanted to play through the finish of the $150,000 tournament Sunday and then begin an extended vacation from professional tennis.
[A003 101] She has complained she lacks enthusiasm for the game.
[A003 102] But she was sidelined with the flu and announced one hour before her scheduled semifinal match with Virginia Wade Saturday night that she would have to forfeit.
[A003 103] Wade advanced to play Tracy Austin in the finals Sunday.

[A003 104] Austin defeated Andrea Jaeger 6-2, 6-2 Saturday night to advance in the tournament.
[A003 105] "*I*'m sorry *I* can not be here tonight to compete in *my* scheduled semifinal match," *Lloyd* said in a prepared statement.

<8> him in [A004 88]   initial-ADV

[A004 83] *Foucault* started his major league career with the Texas Rangers in 1973, setting a club record 12 saves in 69 relief appearances the next year.
[A004 84] In 1977, he was traded to Detroit, where he posted 13 saves.
[A004 85] But the Tigers waived him in 1978.
[A004 86] He was picked up by Kansas City and cut within a month.
[A004 87] He flunked a tryout with Seattle last spring, then the Houston Astros assigned him to their AAA farm team at Charleston, S.C.
[A004 88] When the Charlies cut *him* a month later, *Foucault* said he started looking for a " real job . "

<9> *his* in [A007 7]   initial-ADV

[A007 3] **Bowman**, now coach-general manager of the Buffalo Sabres, will coach the Prince of Wales Conference stars Tuesday night as they seek their fifth consecutive victory over the Clarence Campbell Conference squad coached this year by Arbour.
[A007 4] This 32nd mid-season contest will be held at the almost-completed Joe Louis Arena.
[A007 5] Attendance is expected to exceed the world record for a live audience at a hockey game, 20,009 that witnessed a Blues ' game in St. Louis in 1973.
[A007 6] Prior to the Wales-Campbell clash, the U.S. and Canadian Olympic teams will meet for a one-period exhibition.
[A007 7] Despite *his* remarkable record of achievement in eight years at Montreal, *Bowman* says he has completed the transition to his new home in Buffalo.

<10> *I* in [A007 105] 1PSP-initial-DS

[A007 103] In a sense, Taylor succeeds **Malavasi**, who was offensive coordinator for a brief period in 1978 until he assumed head coaching duties from fired coach George Allen.
[A007 104] The Rams did not have an offensive coordinator for the 1978 and 1979 seasons.
[A007 105] "*I* was trying to do both jobs last year," *Malavasi* said," and this move will free me to concentrate on my duties as head coach.

<11> *his* in [A009 84]   initial-ADV

[A009 82] Human Service Secretary Charles F. Mahoney said Monday that **Gov. Edward J. King** would ask the Legislature this week for $30 million for the mortgage fund.
[A009 83] It was not known immediately what interest rates would be charged hospitals participating in the program, but Mahoney said rates would be lower than current market rates.
[A009 84] In *his* budget message last month, *King* said large mental institutions, such as Danvers and Boston State, cost the state $35,000 a year per patient.

<12> *his* in [A009 98]   initial-ADV

[A009 97] **Thomas,** now an assistant coach at Arizona State University, beat out Renaldo Nehemiah, the world record holder in the high hurdles, and 15-year-old Cynthia Woodhead, who set two world and nine American swimming records in 1979, in the voting for the 50th annual Sullivan Award.
[A009 98] In addition to *his* own personal achievements, *Thomas* said he sees the emergence of American gymnasts as serious contenders in international competition as a source of "a lot of satisfaction."

<13> his in [A010 76]   initial-ADV

[A010 75] **Davidson** isn't.
[A010 76] While still in *his* early 20s, working on the fringes of the Washington power structure and possessed of " a sense one has that things were being screwed up," *Davidson* sought to donate $25 to a taxpayer group.

<14> *he* in [A012 35]   initial-ADV

[A012 33] "It is a heightened expression of friendship toward the American people," said **Norman Forer,** a University of Kansas professor leading the delegation organized by the Committee for American-Iranian Crisis Resolution.
[A012 34] "We congratulate the students for their bold and courageous effort," he added, speaking at a Kennedy Airport news conference just before their departure.
[A012 35] Although *he* expressed concern for the 50 Americans at the U.S. Embassy in Iran and called their detention an " illegal act." *Forer* said the militants ' actions must be weighed " side by side with the anquish of the Iranian people " over what he called the brutal regime under the deposed shah.

<15> *he* in [A013 101]   initial-ADV

[A013 98] **Potts,** who says he is in constant pain because of the police bullet lodged near his spine, last fall suspended all efforts to appeal his sentence and fired his lawyers.
[A013 99] Last week, he was resentenced to die on Feb. 15.

[A013 100] He has rejected appeals to federal courts that could delay his execution for two years.

[A013 101] Since *he* was placed at the state's maximum security prison near Reidsville, *Potts* has complained of pain and requested an operation to remove the bullet, but prison officials say an operation would be too dangerous and probably useless.

<16> *its* in [A016 24]   Other type

[A016 17] Charged with involuntary homicide and injury were Paul Berty, president of the Morhange company, distributor of the product ; Paul Maillard, president of Setico, manufacturer of the powder ; Andre Brunet, technical director of Setico ; Hubert Flahault, president of **Givaudan-France**, manufacturer of the hexachlorophene ; and two Setico employees, Laurence Deroulez and Bernard Civil.

[A016 18] Civil was acquitted.

[A016 19] The other defendants in the two-month trial said they would appeal their sentences, which ranged from one to 20 months in prison, all suspended.

[A016 20] The court also ordered Setico to pay more than $500,000 in fees charged by experts during the long legal process, plus $152,500 in damages and interest to the 45 plaintiffs and $10,000 to the parents of each of the children who died.

[A016 21] More than half of the families of the victims settled out of court before the trial began and received a total of about $1.8 million.

[A016 22] The deaths or paralysis of the babies in various parts of France were linked to a shipment of talcum powder that contained 6 percent hexachlorophene.

[A016 23] Shortly after the deaths, the government put controls on all beauty and toilet products with more than 1 percent hexachlorophene.

[A016 24] In a brief statement issued by *its* head office in Geneva, Switzerland, *Givaudan* management said it " challenged the accusations " made against the company and that its lawyers would appeal those court decisions against it.

<17> *his* in [A016 35]   initial-ADV

[A016 34] He called again and Police Officer David **Cheshire** went to Dillard's home.

[A016 35] Putting *his* ear next to Dillard's head, *Cheshire* heard the music also.

<18> *he* <19> *he* in [A017 54]   initial-ADV (2)

[A017 51] **Belushi** told the audience that he considered the award special because it had been given by students.

[A017 52] "Anyway, I had nothing better to do today," he said.

[A017 53] He added," I'm moved by this, but not a lot . "

[A017 54] When asked if *he* would ever return to " Saturday Night Live," the television show where *he* first gained popularity, *Belushi* said he would not.

<20> *his* in [A017 55]   initial-ADV

[A017 54] When asked if he would ever return to " Saturday Night Live," the
         television show where he first gained popularity, **Belushi** said he would
         not.
[A017 55] When the audience groaned at *his* response, *Belushi* replied," Well maybe
         I will return - when hell freezes over . "

<21> his in [A017 60]   initial-ADV

[A017 59] **Silverheels**, 62, was suffering from a " combination of different illnesses,"
         including cardiac problems, said Jack Staggs, executive director of the
         Woodland Hills hospital.
[A017 60] Hospitalized at the facility last year during *his* recovery from a stroke,
         Silverheels re-entered the hospital Jan. 25.

<22> *its* in [A017 90]   initial-ADV

[A017 87] Public Safety Commissioner Warren Cone on Monday dismissed all
         charges against suspended state police Sgt. David Reed, and denied he had
         tampered with **a disciplinary panel**'s investigation.
[A017 88] (*missing)
[A017 89] Cone said his decision to reinstate Reed was based solely on evidence
         presented during the investigation.
[A017 90] In *its* recommendation that the charges against Reed be dropped, *the panel*
         accused Cone of tampering with its investigation.

<23> *its* in [A017 94]   initial-ADV

[A017 90] In its recommendation that the charges against Reed be dropped, the panel
         accused Cone of tampering with its investigation.
[A017 91] It said Cone should disqualify himself from the case because he had
         become " personally involved . "
[A017 92] Under state police statutes, a three-member panel investigates charges of
         wrongdoing, then makes a recommendation to the commissioner.
[A017 93] Cone said he would not disqualify himself from the case because it was his
         " duty " as commissioner to make the final decision.
[A017 94] In *its* report, *the panel* said Cone presented one of its members with
         damaging evidence that had not been presented by the prosecution.

<24> *I* in [A020 35]   1PSP-initial-DS

[A020 34] Following Tsyganov were Weirather in 1:46.83 ; Ken Read of Canada,
         1:47.41 ; Karl Anderson of Greene, Maine, 1:47.83 ; Herbert Plank of Italy,
         1:47.91 ; Grissmann, 1:47.93 ; Andreas Wenzel of Liechtenstein, 1:47.96,
         and **Walcher**, 1:48.05.
[A020 35] "*I*'m already on the team," *Walcher* said confidently.

<u><25> *his* in [A21 6]</u>   initial-ADV

[A021 5] It was generally welcomed Tuesday by the nuclear industry, environmental groups and congressional Democrats, but Republicans complained **Carter** was moving too slowly.

[A021 6] In *his* message to Congress, *Carter* said radioactive wastes have been generated for more than 30 years," yet past governmental efforts to manage radioactive wastes have not been technically adequate . "

<u><26> his in [A023 68]</u>   initial-ADV

[A023 67] The decision, on a 314-0 vote, to look into the clerk's actions came after a somewhat convoluted series of actions that saw **a Democrat** preempt the Republican-sponsored motion by offering it as his own and Republicans then voting as a bloc to prevent a vote on their own proposal.

[A023 68] By offering the motion as *his* own, *Rep. Richard Bolling, D-Mo.*, was able to cut off debate before Republicans were ready.

<u><27> he in [A024 10]</u> initial-ADV

[A024 8] **Johnson** scored nine straight points early in the second half to stretch out a 54-36 Tennessee lead with 12:40 remaining.

[A024 9] The Bulldogs could get no closer than nine.

[A024 10] When *he* was benched with 10 minutes still left in the game, *Johnson* had a college career total of 1,982 points, 20 above King's three-year total.

<u><28> I <29> my in [A025 25]</u>   1PSP-initial-DS (2)

[A025 23] **Hsu**'s action is still pending in Switzerland.

[A025 24] He said Lord Killanin, the IOC president, has asked him to drop the case.

[A025 25] "*I* must talk to *my* lawyers," *Hsu* said.

<u><30> his in [A025 36]</u>   initial-ADV

[A025 35] Persons who work at educational collaboratives, where handicapped children are instructed, are entitled to the same collective bargaining rights as other public school teachers, **a state Labor Relations Commission official** has ruled.

[A025 36] In *his* ruling, *hearing officer Stuart A. Kaufman* cleared the way for the Massachusetts Federation of Teachers to establish unions at the Shore Collaborative, which serves several of Boston's northern suburbs, and the South Shore Educational Collaborative, stretching from Braintree to Scituate.

<31> *they* in [A026 10]   Other type

[A026 9] Bergmoser had been asked about reports from Tokyo that Douglas A. Fraser, president of the United Auto Workers union, had asked **Mitsubishi** to take part in a joint venture with **Chrysler**.

[A026 10] On a trip to Japan to try to persuade the makers of Toyota and Datsun to build cars in the United States, Fraser told the Detroit News by telephone," Given the relationship *they* already have, I think it would be better for *Mitsubishi and Chrysler* to have a joint venture . "

<32> *his* in [A026 41]   initial-ADV

[A026 38] Some senior IOC members told the newspaper that Killanin has discussed such a trip by telephone with **President Carter** as a way of heading off a threatened U.S. boycott of the Games.

[A026 39] However, the White House would not confirm such a discussion, the newspaper said.

[A026 40] The IOC members told the newspaper a Killanin trip to Moscow could come even before the end of the Winter Olympics now underway in Lake Placid. N.Y.

[A026 41] In *his* news conference Wednesday, *President Carter* dismissed a suggestion raised Monday by his chief counsel, Lloyd N. Cutler, that if the Soviets withdraw their troops from Afghanistan after the Feb. 20 deadline he has set. it would still be possible to send a U.S. team to Moscow.

<33> *he* <34> *his* in [A026 62]   initial-ADV (2)

[A026 58] A probate court judge Wednesday ordered the down payment **Morriseau** had made for Juniper Isle returned because the will of the former owner, the late Sen. Frederick Fayette. has not been settled.

[A026 59] Heirs of the late senator have fought the sale of the 10-acre Lake Champlain island. claiming it could not be sold until ownership is determined.

[A026 60] One of Fayette's brothers claims he owns half of the island.

[A026 61] Sen. Fayette died in 1974 leaving a cash-poor estate and no will.

[A026 62] After learning *he* would get *his* money back, *Morriseau* said he would review the judge's ruling.

<35> *I* in [A027 56]   1PSP-initial-DS

[A027 49] **Heiden** raced in the first pair. so was the first to surpass the previous Olympic record of 2:16.58. set in 1976 by the Soviet Union's Galina Stepanskaya.

[A027 50] Borckink was competing in her second Olympics but was not considered likely to finish even the top 10.

[A027 51] She had injured an ankle last summer and had to wear a cast, causing her a late start here in training for the Olympics.

[A027 52] "Maybe that is our secret," said Dick de Vroomen, the team leader.

[A027 53] Borckink, who understands English but does not speak it much, said she hadn't had many expectations in today's race, but when Visser clocked such a good time she thought," Maybe we have something good going here . "

[A027 54] In January's World Championships she finished only ninth in the 1,500 and her country's Olympic committee told her at first that she was not good enough to come to Lake Placid.

[A027 55] But later in the month she clocked 2:11.35 in a 1,500-meter race in Switzerland, and the decision was reversed.

[A027 56] "*I'*m really surprised at the Dutch girls ' results," said *Heiden*.

<u><36> *his* in [A028 50]</u>   initial-ADV

[A028 48] **Janssen** won three Emmy nominations for his role in " The Fugitive," and his movie credits included " The Shoes of the Fisherman," " The Green Berets." " Two Minute Warning " and " Marooned . "

[A028 49] He returned to television in the 1970s in " O'Hara, United States Treasury " and " Harry O . "

[A028 50] In addition to *his* wife, *Janssen* is survived by his father, Harold Myer of Mesa, Ariz., and his mother, Bernice Janssen of Los Angeles.

<u><37> *I* in [A028 81]</u>   1PSP-initial-DS

[A028 77] The lawsuit results from problems in Crane's campaign headquarters last year that caused mass resignations among his aides and sent the fund-raiser, direct mail expert **Richard A. Viguerie**, packing off to the presidential campaign of former Texas Gov. John Connally.

[A028 78] The suit filed Tuesday not only asks a District of Columbia Superior Court to force Crane to use federal matching funds to pay the bills within 12 days, but it predicts he will fail to get 10 percent of the vote in the Feb. 26 New Hampshire primary.

[A028 79] The court should rule quickly, the suit says, before Crane falls short of 10 percent in two Republican primaries and, thus, becomes ineligible for more federal matching money.

[A028 80] If that happens, the suit says, the campaign probably will not have the cash to pay the bills.

[A028 81] "*I* have no feelings of animosity toward Phil Crane, no rancor," *Viguerie* said in a telephone interview Thursday.

<u><38> *we* in [A029 43]</u>   Other type (1PPP)

[A029 41] "Playing **North Carolina** and playing in the Garden helped us get up for the game," said Rutgers Coach Tom Young.

[A029 42] "But we're still a young team and our youth hurt us down the stretch."

[A029 43] "I thought *we* had the game won in regulation," said *North Carolina* Coach Dean Smith.

<39> *we* in [A029 45]   Other type (1PPP)

[A029 43] "I thought we had the game won in regulation," said **North Carolina** Coach Dean Smith.

[A029 44] "We played good defense at the end, and I know that Rutgers didn't get the shot they wanted . "

[A029 45] "Down the stretch, *we* didn't play like a normal *Carolina* team plays," added Smith.

<40> us in [A029 41] Other type (1PPP)

[A029 37] Overtime play opened with a basket by **Rutgers'** Kevin Black, but Mike Pepper's jump shot 17 seconds later gave North Carolina a 69-68 lead.

[A029 38] Dave Colescott opened it up for North Carolina with a field goal with 2:35 left on the clock.

[A029 39] Jeff Wolf scored North Carolina's last field goal on a dunk shot with 1:15 remaining.

[A029 40] Al Wood wound up with 18 points for North Carolina, which is now 18-5, and Colescott contributed 14.

[A029 41] "Playing North Carolina and playing in the Garden helped *us* get up for the game," said *Rutgers* Coach Tom Young.

<41> *I* in [A030 7] <42> *I* in [A030 8]   1PSP-initial-DS (2)

[A030 5] Ramirez and **Lara** arrived last Wednesday from Palm Springs, Calif., where they played in a Grand Prix tournament.

[A030 6] Ramirez withdrew from that tournament after a slight back injury caused by a distended ligament but was reported in satisfactory condition by the team's physician.

[A030 7] "*I* think we can win.

[A030 8] It is going to be very difficult but *I* think we can make it," said *Lara*.

<43> its in [A032 99]   initial-ADV

[A032 96] Separately, **the Federal Reserve Board** reported that output of the nation's industry increased 0.3 percent in January, the most in four months.

[A032 97] While it was not large, it showed that the performance of the economy continued to defy recession forecasts.

[A032 98] The gain would have been larger if it hadn't been for an 11 percent decline in auto assemblies to an annual rate of 6 million units, down from 6.8 million in December.

[A032 99] In explaining the decision to increase *its* discount rate, *the Federal Reserve Board* made clear in a statement that it felt it had no choice.

<44> *his* in [A033 43]   initial-ADV

[A033 42] **Reynolds**, 37, a former Roman Catholic priest, wrote the comment on his response to a questionnaire distributed last month to all 187 state legislators last month by United Press International.

[A033 43] In *his* apology Friday, **Reynolds** wrote," Many people have said many things to me about welfare.

<45> *her* in [A033 116]   initial-ADV

[A033 110] Guenther's spill and **Zozulya**'s mastery of her competition paved the way for the best finish by other Europeans in the luge since it was included in the Games in 1964.
[A033 111] Residents of East Germany have won the top spot in men's and women's singles and in men's doubles in three of the four past Olympics.
[A033 112] Haspinger, who has chased Guenther for three days, ran through the concrete-lined trough in 43.59 seconds.
[A033 113] Glass, the only East German left in the running, moved into second with a 43.92 run and 2:11.31 overall mark.
[A033 114] Paul Hildgartner of Italy and Anton Winkler of West Germany moved up to third and fourth, respectively, with combined times of 2:11.36 and 2:12.24.
[A033 115] Franz Wilhelmer of Austria had a sparkling run of 44.16, to jump from 10th to fifth, at 2:13.10.
[A033 116] After *her* run of 39.27 seconds, **Zozulya** led Sollmann by four-fifths of a second in combined times, 1:57.42 to 1:58.289.

<46> *his* in [A034 35]   initial-ADV

[A034 33] East German teammates Bernhard Germeshausen and Meinhard Nehmer trailed **Schaerer** by about one-half second, Hans Hilterbrand of Switzerland was fourth, Americans Rushlaw and Howard Siler ranked fifth and sixth and Austrians Franz Paulweber and Fritz Sperling rounded out a four-nation lock on the top eight spots.
[A034 34] The 39-year-old Nehmer won two gold medals in the 1976 Olympics with Germeshausen as his brakeman.
[A034 35] To achieve *his* hopes of a second good run, **Schaerer** has to avoid the hopes of Rushlaw.

<47> *he* in [A034 102]   initial-ADV

[A034 101] In some countries, the best thing is to be a ski racer," **Kiesel** said Friday.
[A034 102] Although *he* didn't mention any countries by name, **Kiesel** clearly was referring to the Russian team which had just overwhelmed the United States for the second day in a row at the XIII Winter Olympics.

<48> *it* in [A037 70]   Other type

[A037 66] Aikens, a first baseman who underwent surgery on **his left knee** after an injury in September, said the knee was at about 90 percent efficiency.
[A037 67] "I was pleasantly surprised," Manager Jim Frey said at the end of a two-hour workout for 22 players who showed up for the start of a special rehabilitation and instructional camp at Terry Park.

[A037 68] "I was expecting to see him limping and favoring it," Frey said.
[A037 69] "But he wasn't at all.
[A037 70] At the end of the day, *it (Aikens ' knee)* looked a little puffy.

<u><49> *I* in [A037 68]</u>   1PSP-initial-DS

[A037 67] "I was pleasantly surprised," **Manager Jim Frey** said at the end of a two-
    hour workout for 22 players who showed up for the start of a special
    rehabilitation and instructional camp at Terry Park.
[A037 68] "*I* was expecting to see him limping and favoring it," *Frey* said.

<u><50> *I* in [A037 90]</u>   1PSP-initial-DS

[A037 89] Managers for veteran Fel Clemente and novice **Rocky Lockridge** each
    predicted victory in their featherweight boxing match tonight at the Ice
    World.
[A037 90] "*I*'m not predicting anything," said *Lockridge*, 10-0, a freshman at William
    Paterson College.

<u><51> *I* in [A037 112]</u> 1PSP-initial-DS

[A037 108] **Duva**, for one, feels Thompson will find out differently.
[A037 109] "This is a major stepping stone for Rocky.
[A037 110] If he wins this fight, Top Rank is looking for him to step into the TV
    picture.
[A037 111] They know Rocky's the newest star on the horizon . "
[A037 112] "But *I*'m not so much worried about the title fight," *Duva* added.

<u><52> *we* in [A038 13]</u>   Other type (1PPP)

[A038 12] Detroit defenseman Reed Larson feels the Red Wings are about to explode
    and **Los Angeles Kings** ' Coach Bob Berry agrees with him.
[A038 13] "Sure. *we* worked hard for this victory." Berry said following *the Kings* ' 4-
    2 National Hockey League triumph Monday night.

<u><53> *we* in [A038 71] <54> *we* in [A038 72]</u>   Other type (1PPP) (2)

[A038 65] In fact. the circus comprises only a few of the thousands of models
    Wagenfuehr has built - and still builds.
[A038 66] The painted models - some wood. others plastic - overflow showcases and
    hang from the ceiling of his downtown barber shop.
[A038 67] Ships. cars, planes, soldiers and dinosaurs are everywhere.
[A038 68] A customer leaning back in a barber chair for a shave is confronted by a
    model of a Navy zeppelin hanging above.
[A038 69] His house next door brims with more models, a doll collection and other
    toys. and the rock jewelry he now loves to make.
[A038 70] Once he built a tiny replica of his church in New Braunfels, complete with
    stone walls.

[A038 71] "*We*'ll never grow old this way.

[A038 72] *We*'ll never grow up," said *Wagenfuehr*'s patient wife of 18 years, *Eleanor*.

<55> *I* <56> *I* in [A038 74]   1PSP-initial-DS (2)

[A038 73] They have no children - just **Wagenfuehr**.

[A038 74] "My, my, *I* don't know how many models *I*'ve built," said *Wagenfuehr*.

<57> *I* <58> *I* in [A041 25]   1PSP-initial-DS (2)

[A041 22] **Malmquist**'s final point total was 391.915.

[A041 23] Although Wehling only managed a ninth-place finish in the 15-km race, it was enough to give him his third gold medal in the Nordic Combined event.

[A041 24] His point total was 432.200.

[A041 25] "*I* wish *I* could be like Bill Koch.

[A041 26] He has so much intestinal fortitude," said *Malmquist*, referring to the top U.S. cross-country skier.

<59> *their* in [A043 82]   initial-ADV

[A043 81] "Passenger car frames and structural components have been a major part of this company's business since early in this century, and we will continue to identify and demonstrate the advantages of isolated, separate frames to **the auto makers** whenever we can," A. O. Smith president J. R. Parker said.

[A043 82] "In some of *their* new passenger cars, *the manufacturers* will be using all-new types of chassis structural components, and A. O. Smith expects to be an important supplier of these products as well," Parker said.

<60> *her* in [A044 71]   Other type

[A044 70] Jaeger defeated **Barker** in straight sets earlier this month at a tournament in Seattle, the first and only time they have squared off.

[A044 71] "I'm really keen to play *her* again," *Barker* said of her meeting with *the high school freshman from Lincolnshire, Ill.*

<61> *his* in [A045 39]   initial-ADV

[A045 36] The officers said they found **King** sprawled partially nude on the bed in a semiconscious state.

[A045 37] The preliminary hearing Wednesday lasted several hours, during which the woman testified.

[A045 38] The sodomy charges are second-degree felonies punishable by a possible 1-15 years in prison and the sexual abuse charges are third-degree felonies which are punishable by a prison term of up to five years.

[A045 39] Since *his* sophomore year at Tennessee, *King* has had several brushes with the law.

<u><62> *our* in [A045 100]</u>   Other type (1PPP)

[A045 99] After John Adams ' layup for **State**, Derrick Hord hit from eight feet and
      Macy drove for a layup and a 67-57 lead - Kentucky's biggest of the game -
      with 4:18 remaining.
[A045 100] "I was pleased with *our* effort," said *State* coach Jim Hatfield, whose
      team fell to 12-13 overall and 6-11 in the SEC.

<u><63> *our* <64> *us* in [A046 6]</u>   Other type (1PPP) (2)

[A046 5] But until Wednesday, **the court** had never approved direct payment to non-
      public schools without neutral organizations serving as buffers to ensure
      the constitutionally required separation of church and state.
[A046 6] "None of *our* (previous) cases requires *us* to invalidate these
      reimbursements simply because they involve payments in cash," Justice
      Byron R. White wrote for *the court* Wednesday.

<u><65> *his* in [A046 33]</u>   initial-ADV

[A046 32] Mansfield also said he thought the recent visit to Japan by **United Auto
      Workers President Douglas A. Fraser** had greatly helped " both sides in
      understanding their respective situations . "
[A046 33] During *his* visit, *Fraser* urged Japanese automakers to curb exports and
      start production in the United States. warning they could otherwise face
      congressional moves to protect America's beleaguered auto industry.

<u><66> *I* in [A046 51]</u>   1PSP-initial-DS

[A046 50] **Assalone** said several other states are considering such proposals.
[A046 51] "*I* think it's important to keep the national momentum going on this thing.
[A046 52] It's not so important what the effect will be in Rhode Island but that it will
      spread to other states too." *he* said . (0)

<u><67> *I* in [A046 42]</u>   1PSP-initial-DS

[A046 41] **A state lawmaker** says he has 50 House co-signers on a proposal he plans
      to file next week that would prohibit the major oil companies from
      operating service stations in Rhode Island.
[A046 42] "*I* think they're one of the few industries that enjoys a total monopoly from
      the ground to their pocketbook." said **Rep. John A. Assalone, D-
      Coventry**.

<u><68> *I* in [A046 68]</u>   1PSP-initial-DS

[A046 66] "He was picking up the money for **Jenrette**, period.
[A046 67] He was a courier."
[A046 68] "*I* absolutely deny that," *Jenrette* said.

<u><69> *I* <70> *I* in [A046 69]</u>   1PSP-initial-DS (2)

[A046 68] "I absolutely deny that," Jenrette said.
[A046 69] "*I* know what he did and what *I* did," *Jenrette* told reporters in
        Washington on Wednesday.

<u><71> *he* in [A046 97]</u>   initial-ADV

[A046 96] **Kennedy** made the call Wednesday in a speech on energy at the Franklin
        Pierce Law Center.
[A046 97] Asked specifically what *he* would do about the unfinished plant of Public
        Service Co. at Seabrook, *Kennedy* referred to his call last month for
        converting the plant to coal.

<u><72> *his* in [A047 66]</u>   initial-ADV

[A047 65] In 1966, **Lauck** retired to his Lake Hamilton home near Hot Springs.
[A047 66] Since *his* retirement, **Lauck** had served a five-year stint on the state
        Racing Commission.

<u><73> *they* in [A047 100]</u>   initial-ADV

[A047 94] I don't know if **the Soviets** work on their skating any more than we do on
        our own.
[A047 95] I just think their tactics - having all the players in motion at once - force
        the players to be better skaters at an earlier age.
[A047 96] Then they're supplemented by skating drills," said Brooks.
[A047 97] "You can have all the skating drills in the world, but if you take your good
        skaters and play a different style - a style that doesn't bring the skating out
        in them - it's stupid.
[A047 98] It's not going to gain the dividends you want . "
[A047 99] And now. the metal dividend the Americans want is in sight.
[A047 100] Unless *they*'ve been bluffing the last five games, *the Soviets* seems as
        mortal as any other club. regardless of the fact they've won the last four
        Olympic golds in hockey.

<u><74> *we* <75> we in [A048 10]</u>   Other type (1PPP) (2)

[A048 9] But the **Hoosiers**, led by Mike Woodson and Butch Carter improved their
        accuracy to 70 percent in the final 20 minutes while MSU slumped to 42
        percent.
[A048 10] "I think the most important part of the game was when *we* were down 56-
        50 and *we* scored eight straight points and went ahead," Coach Bobby
        Knight said of the *Hoosier* comeback.

<76> *its* in [A049 77] initial-ADV

[A049 76] **The committee** last week obtained a Superior Court order barring the
town from spending a $252,000 budget surplus on anything but school
expenses.
[A049 77] As *its* part of the agreement, *the committee* said it will cut energy costs by
closing schools at 3:15 p.m. each day and shut down all heating systems by
April 30th.

<77> *I* in [A051 80]   1PSP-initial-DS

[A051 79] **Frances Vrhovac**, a friend of Mrs. Kovatch for 34 years, said the woman
became withdrawn and refused to talk about the tombstone.
[A051 80] "*I* saw that it was a traumatic experience for her to see that stone," said
Mrs. Vrhovac, who was with Mrs. Kovatch when she first saw the
inscription.

<78> *her* in [A051 80]   Other type

[A051 79] Frances Vrhovac, a friend of **Mrs. Kovatch** for 34 years, said the woman
became withdrawn and refused to talk about the tombstone.
[A051 80] "I saw that it was a traumatic experience for *her* to see that stone," said
Mrs. Vrhovac, who was with *Mrs. Kovatch* when she first saw the
inscription.

<79> *his* in [A054 108]   initial-ADV

[A054 107] The 29-year-old **Sittler** has been feuding with Toronto president Harold
Ballard and General Manager Punch Imlach for most of the season.
[A054 108] After the Leafs traded *his* best friend, right wing Lanny McDonald,
Sittler resigned his captaincy.

<80> *I* in [A055 3]   1PSP-initial-DS

[A055 2] But **Claudia Giordani** says the problems aren't irreversible.
[A055 3] "*I* think this result today shows we can be good again," said *Giordani* after
all four of the Italian women in the Olympic slalom Saturday finished in the
top 10.

<81> *I* in [A055 19]   1PSP-initial-DS

[A055 17] **Cooper** had the best American finish in the Olympic giant slalom -
seventh - and she followed it up Saturday with another American best - an
eighth-place finish in the slalom.
[A055 18] "It's coming.
[A055 19] *I*'m skiing better and with more confidence each time out," *Cooper* said
Saturday.

<82> *I* in [A055 43] 1PSP-initial-DS

[A055 41] **Kinshofer** had a time of 43.76 seconds, then waited for Wenzel to ski.
[A055 42] "After a very good second run, I thought for a few moments that I could
win the gold.
[A055 43] *I* knew Hanni could have trouble," **Kinshofer** said.

<83> *me* in [A055 33]   1PSP-initial-DS

[A055 32] **Christa Kinshofer of West Germany** won five World Cup giant slalom
races last season, so it was somewhat of a surprise when she won her only
Olympic gold medal, a silver, in the special slalom.
[A055 33] "This makes *me* very happy," *the blonde-haired 19-year-old* said after
finishing second Saturday to Hanni Wenzel of Liechtenstein in the slalom
at the 1980 Winter Olympics.

<84> *I* in [A059 30] 1PSP-initial-DS

[A059 17] "That's a sign of a really outstanding ballclub that they can generate
things." said **Boston College Coach Tom Davis**.
[A059 18] But Roosevelt Bouie hit two free throws and followed it with a dunk, and
Louis Orr hit a three-point play to make the score 34-25 and Syracuse was
on its way back.
[A059 19] "Some games we come out and play well but in most games we have not
been a good first-half team." said Syracuse Coach Jim Boeheim.
[A059 20] "They (Boston College) executed well and got a lot of movement and we
were just standing around " in the first half.
[A059 21] The Eagles hit an amazing 76 percent of their field goal attempts in the
first half when they outrebounded the taller Orangemen 11-7.
[A059 22] But Orr and freshman Erich Santifer led the second-half charge as
Syracuse boosted its record to 24-2.
[A059 23] The victory gave the Orangemen a tie for the Big East title with St. John's
and Georgetown, all with 5-1 records.
[A059 24] Boston College is 18-8 overall and 2-4 in the conference.
[A059 25] "Orr was the key.
[A059 26] He's the key to our team." said Boeheim.
[A059 27] "He had his worst first half of the year.
[A059 28] The second was his best . "
[A059 29] Orr led Syracuse with 23 points and Santifer had 17, including eight
straight during a 10-point spurt by the Orangemen in which they took a 70-
55 lead with 5:02 left in the game.
[A059 30] "*I* thought we gave Syracuse a good run and they showed their poise at the
start of the second half." said *Davis*.

<85> *we* <86> *our* in [A059 58]   Other type (1PPP) (2)

[A059 56] Then **the Jayhawks** had to beat Oklahoma State at Lawrence, which they
finally accomplished in overtime, 84-74.
[A059 57] In the other regular season finale Saturday night, Oklahoma broke a five-
game losing skid by downing Nebraska, 78-60.
[A059 58] "I was very pleased that *we* kept *our* poise tonight," said ***Kansas*** Coach
Ted Owens.

Appendix 4-3-1: Cataphora examples where Pronouns occur in an initial adverbial and the antecedent occurs in the main clause other than in the subject position.

(1) While *he* lived, few Japanese had heard of *Atuhito Nakata*, but for his country he was an important symbol. (*Newsweek*, 10 May '93)
— taken from Kanzaki (1997:104)

(2) Though *he* tried not to think of it[1], at twenty-nine *Tommy Castelli's* life was a screaming bore.
B. Malamud, *The Prison*, 97
— taken from Kanzaki (1997:107)

(3) Scrimshawing took time. And, once *his* ship had reached the whaling grounds, time was something *every whaleman* had a great plenty of.
(Shapiro, *The Story of Yankee Whaling*: 291 (1959))
— taken from Carden (1982:369)

(4) Unless *he's* very hungry and therefore desperate, chances are *a shark* will not attack unless . . .
(*Yachting*, Dec. 75)   — taken from Carden (1982:370)

(5) [A016 24] In a brief statement issued by *its* head office in Geneva , Switzerland, *Givaudan* management said it "challenged the accusations" made against the company and that its lawyers would appeal those court decisions against it.

(6) Despite *its* long reach, the testing activities of *ETS* are not overseen by any public agency.
(*Readers Digest* 3/80) — taken from Van Hoek (1997:116)

(7) Never mind *their* guilt or innocence, the executions of *atomic spies Julius and Ethel Rosenberg* 36 years ago were illegal and represent a breakdown in the criminal justice system.
(*San Diego Union* 8/89) — taken from Van Hoek (1997:117)

(8) Without *his* own dramatic presence, *Andy Warhol's* work remains great art.
(*Los Angeles Herald Examiner* 2/6/89)
— taken from Van Hoek (1997:117)

(9) As *he* walked past City Hall on Wednesday, *attorney David Miller's* eyes were drawn to a banner proclaiming "DRUG FREE THE WAY TO BE."
(*San Diego Union* 11/12/89)
— taken from Van Hoek (1997:117)

---

[1] This 3PP *it* also can be identified as cataphora.

(10) With some in *his* own party skeptical and the whole GOP gleeful, is California
ready for ... **The Rebirth of *Jerry Brown*** ?
(*Los Angeles Herald Examiner* 2/6/ 89, title of article)
— taken from Van Hoek (1997:117)

(11) Though *it* was completed in 1989, you'd like to nominate erstwhile journalist
Michael Moore's ***Roger and Me*** as the first film of the nineties.
(*Tikkun* 11/ 89)
— taken from Van Hoek (1997:119)

(12) By matching *his* parents' characters and backgrounds with the findings of
scholarly studies, they decided that ***Michael Lee Hall*** faced a precise 40% chance
of being abused within the next five years.
(*Los Angeles Times* 2/1/89)   — taken from Van Hoek (1997:119)

(13) In this temple, as in the hearts of the people for whom *he* saved the Union, the
memory of *Abraham Lincoln* is enshrined forever.
source unknown — taken from Van Hoek (1997:120)

(14) Mistaking *him* for a burglar, Barbara shoots ***Carol's boyfriend.***
(*Chicago Tribune* 1/5/91, blurb for a TV program)
— taken from Van Hoek (1997:121)

(15) Once *it*'s built in Simi Valley, California, scholars will flock to ***the Ronald
Reagan Presidential Library*** to sift through state papers.
(*Time* 1/16/ 89) — taken from Van Hoek (1997:121)

(16) Aside from *his* schoolboy experiences, one cannot present ***George Orwell*** as a
victim of child abuse.
(Leonard Shengold, *Soul Murder*, p. 69) — taken from Van Hoek (1997:121)

Appendix 4-3-2: Cataphora examples where a plural 1PP in the initial DS is coreferential with a full NP appearing in the reporting clause.

(1)

[A014 93] "*We* only transported one person to the hospital by ambulance for treatment because of the gas, " said *fire department* spokesman Dale Everett.

(2)

[A044 2] "The cost of funds is climbing so rapidly *we* haven't any other choice," David Reed, senior vice president of *California Federal's loan division*, said Wednesday .

(3)

[A027 52] "Maybe that is *our* secret," said Dick de Vroomen the *team* leader.

(4)

[A029 43] "I thought *we* had the game won in regulation," said *North Carolina* Coach Dean Smith.

(5)

[A029 41] "Playing North Carolina and playing in the Garden helped *us* get up for the game," said *Rutgers* Coach Tom Young.

(6)

[A045 100] "I was pleased with *our* effort, " said *State* coach Jim Hatfield, whose team fell to 12-13 overall and 6-11 in the SEC.

(7)

[A059 58] "I was very pleased that *we* kept *our* poise tonight, " said *Kansas* Coach Ted Owens.

(8)

[A029 45] " Down the stretch, *we* didn't play like a normal *Carolina* team plays, " added Smith.

(9)

[A038 13] " Sure, *we* worked hard for this victory, " Berry said following *the Kings'* 4-2 National Hockey League triumph Monday night.

(10)

[A046 6] " None of *our* (previous) cases requires *us* to invalidate these reimbursements simply because they involve payments in cash, " Justice Byron R. White wrote for *the court* Wednesday.

(11)

[A048 10] " I think the most important part of the game was when *we* were down 56-50 and *we* scored eight straight points and went ahead, " Coach Bobby Knight said of *the Hoosier* comeback.

(12)

[A038 72] *We*'ll never grow up , " said *Wagenfuehr's patient wife of 18 years,*
Eleanor.

Appendix 5-1: Informant test sheets for testing Reinhart's claim on the syntactic difference between initial Verb-phrasal PPs and initial Sentential PPs

Acceptability test:

For each example, please select one from the scale below.

1	2	3	4	5
No problem	More or less OK, but sounds a little bit odd	Sounds odd but not sure if acceptable	Sounds odd and probably unacceptable	Sounds odd and completely unacceptable

Examples:

e.g.1	Kate found the car keys in her letter box.	(OK) 1 2 3 4 5 (Odd) √
e.g.2	Rosa was hit and Max by John.	(OK) 1 2 3 4 5 (Odd) √

Please feel free to give any comments on the examples in blank spaces.

1.	John found the book he needed in the library.	(OK) 1 2 3 4 5 (Odd)
2.	Felix didn't realise until whose remark he is a failure.	(OK) 1 2 3 4 5 (Odd)
3.	With her boss, does she argue?	(OK) 1 2 3 4 5 (Odd)
4.	In Ben's picture of her, what did she find?	(OK) 1 2 3 4 5 (Odd)
5.	In Detroit, who did the gangsters kill with the shotguns they smuggled?	(OK) 1 2 3 4 5 (Odd)
6.	She usually argues with her students about how she marks their essays and exams.	(OK) 1 2 3 4 5 (Odd)
7.	With that customer, stop arguing about the quality and defects of our products!	(OK) 1 2 3 4 5 (Odd)
8.	In front of Felix, who holds a candle?	(OK) 1 2 3 4 5 (Odd)

9.	Stop arguing with your boss!	(OK) 1 2 3 4 5 (Odd)
10.	In this brand-new product, find any defects, if you can!	1(OK) 2 3 4 5(Odd)
11.	With her new boss, why does Rosa argue?	(OK) 1 2 3 4 5 (Odd)
12.	In Ben's picture a scratch, did find Rosa?	(OK) 1 2 3 4 5 (Odd)
13.	With your children, why do you usually argue about their exam results?	(OK) 1 2 3 4 5 (Odd)
14.	In my next picture, find a scratch, if you can!	1(OK) 2 3 4 5(Odd)
15.	Kate argued with her boss yesterday.	(OK) 1 2 3 4 5 (Odd)
16.	In the new version of Windows, did you find any bugs that you haven't come across before?	(OK) 1 2 3 4 5 (Odd)

314

No.	Sentence	Rating
17.	With her boss, she argues.	(OK) 1 2 3 4 5 (Odd)
18.	In front of the demo, who is holding the large picture of Nelson Mandela?	(OK) 1 2 3 4 5 (Odd)
19.	With her students, she usually argues about how she marks their essays and exams.	(OK) 1 2 3 4 5 (Odd)
20.	In Ben's picture, did Rosa find a scratch?	(OK) 1 2 3 4 5 (Odd)
21.	In Detroit, who did the gangsters kill?	(OK) 1 2 3 4 5 (Odd)
22.	Day after day, with her bosses she argued, and with her husband she quarrelled.	(OK) 1 2 3 4 5 (Odd)
23.	With her students, does she usually argue about how she marks their essays and exams?	(OK) 1 2 3 4 5 (Odd)

24.	In the lucky bag Dad gave you, what did you find?	(OK) 1 2 3 4 5 (Odd)
25.	With your boss, stop arguing!	(OK) 1 2 3 4 5 (Odd)

Appendix 5-2-3: AT examples in which the initial-ADV cataphora pattern invites an anaphoric reading

(52-1)

[A006 5] A trial court judge said there was insufficient evidence to deport *<REF=1 him*, but an appeals court reversed the decision - prompting *(1 Fedorenko 1)* to file a Supreme Court appeal recently.

[A006 6] In asking the Supreme Court to be permitted to retain *<REF=1 his* citizenship, *(1 Fedorenko 1)* said <REF=1 he had 30 years of good conduct in the United States.

(52-2)

[A011 59] But *(102 Brown 102)*, who so far has appeared to be running a distant third in the Democratic race, refused to say what percentage of the turnout he will need to do well in the caucuses.

[A011 60] "We 'll know it when we see it , " *<REF=102 he* said .

[A011 61] Asked whether *<REF=102* he would take more votes away from Sen. Edward M. Kennedy than from President Carter, *(102 Brown 102)* at first said , " <REF=102 I do n't think it 's clear at this point .

(52-3)

[A022 16] Over the years , *(1 DeLury 1)* gained a reputation on both sides of the bargaining table as {{1 an effective labor leader and an innovative negotiator 1} .

[A022 17] When *<REF=1 he* retired . *(1 DeLury 1)* was succeeded by the union's secretary-treasurer , Edward Ostrowski, 56 , who said of <REF=1 his death : "We have lost (1 a major figure in our lives 1) "

(52-4)

[A035 87] Asked to comment on various proposals for budget cutting or tax increases, *(104 Reynolds 104)* wrote : "No!

[A035 88] No!

[A035 89] No income tax.

[A035 90] No more taxes.

[A035 91] Limit spending.

[A035 92] Put the Niggers back to work. "

[A035 93] "*<REF=104 I* deeply regret and apologize for *<REF=104 my* wrong choice of the next regular House session. "

[A035 94] "This type of racism can not be allowed nor tolerated by any member of the Connecticut General Assembly. " Milner said in a letter to Abate.

[A035 95] Abate said a final decision on censure would have to be made by the full House.

[A035 96] He said he was having a legal opinion researched on the proper course for censure in anticipation that a formal move would be made in that direction.

[A035 97] State Sen. Sanford Cloud Jr. a black representing a Hartford district, said, "I am appalled.

[A035 98] I question *<REF=104 his* ability to be a state representative."

[A035 99] In *<REF=104 his* Friday apology. *(104 Reynolds 104)* wrote , " Many people have said many things to <REF=104 me about welfare .

(52-5)

[A039 87] " <REF=93 I started painting these trains after that, " said *(93 Nelson 93)*, who has had a couple of exhibitions.

[A039 88] " *<REF=93 I* also like to paint wildlife. "

[A039 89] <REF=93 He has n't painted much for the past couple of years, however.

[A039 90] "*<REF=93 I* do n't know where *<REF=93 I* 'd put another picture, " *<REF=93 he* said, grinning.

[A039 91] " *<REF=93 I* 've got all of *<REF=93 my* walls covered."

[A039 92] In addition to *<REF=93 his* steam engines and wild animals , *(93 Nelson 93)* also has done several portraits of friends and family members .

(52-6)

[A044 31] Until a few years ago when <REF=30 her health began fading, Washingtonians could always spot the hottest party in town when they saw *(30 Mrs. Longworth 30)* 's vintage black Cadillac parked out front of some notable 's house.

[A044 32] And <REF=30 she was often seen browsing in a local bookstore wearing the broadbrimmed hat that was <REF=30 her trademark.

[A044 33] <REF=30 She loved to play poker and once said <REF=30 she made $15,000 one year "which is damn good. "

[A044 34] <REF=30 She knew every president since Benjamin Harrison who was in office from 1889 until 1893.

[A044 35] And whether <REF=30 she liked them or not, <REF=30 she rarely hesitated to say exactly how <REF=30 she felt about them.

[A044 36] <REF=30 She was a favorite of Harry Truman, John Kennedy, Lyndon Johnson and Richard Nixon.

[A044 37] <REF=30 She was not on particularly good terms with Warren Harding and Woodrow Wilson.

[A044 38] <REF=30 She once said Dwight Eisenhower bored <REF=30 her.

[A044 39] Candid and uncontrived comments often spiced *<REF=30 her* deliciously naughty conversations. making *<REF=30 her* {{30 a sought-after guest in (35 the capital 35) 's social circuit 30}.

[A044 40] And *<REF=30 she* never lost the rapier wit that prompted one of *<REF=30 her* more famous remarks : " If you have n't got anything nice to say about anybody , come . sit next to *<REF=30 me.* "

[A044 41] During an interview on *<REF=30 her* 90th birthday when many of Washington's VIPs made a pilgrimage to *<REF=30 her* house for champagne and rum cake . *(30 Mrs. Longworth 30)* said , " *<REF=30 I* do n't think *<REF=30 I* 'm insensitive or cruel .

(52-7)

[A044 63] " Tell your wretched people *<REF=30 I* send them *<REF=30 my* curses, " *(30 Mrs. Longworth 30)* replied.

[A044 64] Then *<REF=30 she* relented .

[A044 65] Asked at that time how *<REF=30 she* viewed life as *<REF=30 she*
approached *<REF=30 her* 90th year , *(30 Mrs. Longworth 30)* replied , "
What do you mean , View life ? '

(52-8)
[A060 34] *(28 Cannon 28)* told a few more jokes and then called attention to a small
bandage alongside *<REF=28 his* right eye.
[A060 35] The result of minor surgery, *<REF=28 he* said.
[A060 36] But *<REF=28 he* quickly added : " *<REF=28 I* 'm reluctant to say that
<REF=28 I had minor surgery , for fear *<REF=28 I* will see on television
tonight that *<REF=28 I* had a lobotomy over the weekend . "
[A060 37] Winding up *<REF=28 his* monologue , *(28 Cannon 28)* said : " If anyone
is offended by any of the jokes *<REF=28 I* have told, *<REF=28 I* want you
to know that *<REF=28 I* am like Ronald Reagan."

Bibliography

Aoun, J.	1986	*Generalised Binding*, Dordrecht, Foris.
Ariel, M.	1985a	Givenness Markig, PhD thesis, Tel-Aviv University
Ariel, M.	1990	*Accessing Noun Phrase Antecedents*. New York: Routledge.
Biller-Lappin, Y.	1983	*Backwards Anaphora in Discourse*, MA Thesis, Tel-Aviv University.
Bolinger, D.	1977	*Pronouns and Repeated Nouns*. Bloomington, Indiana: Indiana University Linguistics Club.
Bosch, P.	1983	*Agreement and Anaphora: A Study of The Role of Pronouns in Syntax and Discourse*. New York: Academic Press.
Brennan, S., Fridman, M. and Pollard, C.	1987	A Centering Approach to Pronouns. *Proceedings of The 25th Annual Meeting of The ACL(ACL '87)*, pp.155-162. Stanford, CA, USA.
Brown, G. and Yule, G.	1983	*Discourse Analysis*, Cambridge University Press.
Bühler, C.	1982	The Deictic Field of Language and Deictic Words, in Javella, R. and Klein (Eds.), W. *Speech, Place and Action: Studies in Deixis and Related Topics* (pp 9-30), New York, Wiley
Carden	1982	Backwards Anaphora in Discourse Context. *Journal of Linguistics* 18: 361-387.
Carden	1986a	Blocked Forwards Coreference: Theoretical Implications of The Acquisition Data. in B. Lust (Ed.), *Studies in The Acquisition of Anaphora. Vol. 1*, pp.319-357. Reidel Publishing.
Carden	1986b	Blocked Forwards Coreference: Unblocked Forwards Anaphora: Evidence For An Abstract Model of Coreference. in Farley, A.M. et al. *Papers From The General Session at The 22nd Regional Meeting*, Chicago Linguistic Society
Chafe, W.L.	1970	*Meaning and The Structure of Language*, University of Chicago Press.
Chafe, W.L.	1974	*Language and Consciousness*. Language, 50 (I) pp.111-133
Chomsky, N.	1965	*Aspects of The Theory of Syntax*. Cambridge, Mass.: MIT Press.
Chomsky, N.	1980	*Rules and Representations*, New York, Columbia University Press.
Chomsky, N.	1981	*Lectures on Government and Binding*. Dordrecht: Foris.
Chomsky, N.	1985	*Knowledge of Language: Its Structure, Origin and Use*, New York, Praeger.
Chomsky, N.	1986	*Barriers*, Cambridge, Mass. MIT Press.

Cornish, F.	1986a	Anaphora and Semantic Interpretation: Review Of: T. Reinhart. Anaphora and Semantic Interpretation, *Lingua:-International-Review-Of-General-Linguistics, Amsterdam, Netherlands (Lingua). 1986 Mar., 68:2-3, 241-256*
Cornish, F.	1986b	*Anaphoric Relations in English and French: A Discourse Perspective.* London and Canberra: Croom-Helm
Cornish, F.	1996	'Antecedentless' Anaphors: Deixis. Anaphora. Or What? Some Evidence From English and French. *Journal of Linguistics,* 32:19-41
Deemter, C. J. van	1989	*Forward References in English, IPO-Annual-Progress-Report,* Eindhoven, Netherlands (IPOAPR). 1989, 24, 113-23
Delisle, G. A.	1973	*Discourse and Backward Pronominalization .* Bloomington, Indiana: Indiana University Linguistics Club.
Emmott, C.	1997	Narrative Comprehension: A Discourse Perspective, Clarendon Press. Oxford
Fligelstone, S.D.	1989	Enriched Annotation of The Skeleton Treebank For Cohesion Resolution. Internal UCREL Report. UCREL, University of Lancaster
Fligelstone, S.D.	1990a	Supplementary Guidelines For Anaphora Annotation, Internal UCREL Report. UCREL, Lancaster University Lancaster
Fligelstone, S.D.	1990b	Developing A Scheme For Annotation)Text to Show Anaphoric Relations. in *Proceedings of The 11th International Conference on English Language Research on Computer Corpora.* Berlin 1990 (Edited By G. Leitner). Berlin: Mouton De Gruyter.
Fligelstone, S.D.	1991	A Description of The Conventions Used in The Lancaster Anaphoric Treebank Scheme (Second Edition). Internal UCREL Report. University of Lancaster
Garside, R.	1993	The Marking of Cohesive Relationship: Tools For The Construction of A Large Bank of Anaphoric Data. *ICAME Journal* 17. pp.5-27.
Garside, R. and McEnery, A.	1993	Treebanking: The Compilation of A Corpus of Skeleton Parsed Sentences. in Black. E. Garside. R. and Leech, G. (Eds) *Statistically-Driven Computer Grammars of English: The IBM Lancaster Approach.* Amsterdam: Rodopi. 1993:17-35
Garside, R., Fligelstone. S. and Botley. S.	1997	Discourse Annotation: Anaphoric Relations in Corpora. in Garside. R.. Leech G. and Mcenery, 1997.
Garside, R., Leech G. and McEnery, A. (eds.)	1997	*Corpus Annotation: Linguistic Information From Computer Text Corpora.* London. Longman

Garside, R., Leech G. and Sampson G. (eds.)	1987	*The Computational Analysis of English: A Corpus-Based Approach.* London, Longman
Givón T.	1990	*Syntax: A Functional Typological Introduction,* John Benjamins, Amsterdam.
Givón, T.	1993	*English Grammar: A Function-Based Introduction,* John Benjamins, Amsterdam.
Greene, S.B., Gerrig, J.G., McKoon, G., Ratcliff, R.	1994	Unheralded Pronouns and Management By Common Ground, *Journal of Memory and Language* 33:511-526
Gundel,.J.K.; Hedberg, N. and Zacharski, R.	1993	Cognitive Status and The Form of Referring Expressions in Discourse. *Language* . 274-307.
Haegeman, L.	1994	*Introduction to Government and Binding Theory (2nd Edn),* Blackwell.
Haegeman, L. and Guéron, J.	1999	*English Grammar: A Generative Perspective,* Blackwell
Halliday, M.A.K. and Hasan, R.	1976	*Cohesion in English,* London: Longman.
Horrocks, G.	1987	*Generative Grammar,* London: Longman.
Jackendoff, R.	1972	*Semantic Interpretation in Generative Grammar,* Cambridge, Mass. MIT Press
Jones, P.	1994	The Great Anaphora Hoax: A Radical Critique of Syntax and Pseudo-Pragmatic Approaches to Coreference Phenomena. *Paper Given at PALA Conference,* Sheffield Hallam University.
Just, M.A. and Carpenter, P.A.	1980	A Theory of Reading: From Eye Fixations to Comprehension. *Psychological Review,* 87, pp. 329-354.
Kamp, H. and Reyle, U.	1993	*From Discourse to Logic,* Dordrecht, Kluwer.
Kanzaki, T.	1997	*Nichi-Ei-Go Daimeishi No Kenkyu (Pronouns in Japanese and English),* Kenkyusha Press, Tokyo
Kuno, S.	1972	Functional Sentence Perspective: A Case Study From Japanese and English. *Linguistic Inquiry* 3: 269-320.
Kuno, S.	1975	*Three Perspectives in The Functional Approach to Syntax.* in R. E. Grossman, L. J. San and T. J. Vance (Eds.) *Papers From The Parasession on Functionalism. Chicago:* Chicago Linguistic Society. 276-336.
Kuno, S.	1987	*Functional Syntax: Anaphora, Discourse, and Empathy,* University of Chicago Press.

Kuno, S. and Takami, K.	1993	*Grammar and Discourse Principles: Functional Syntax and GB Theory,* University of Chicago Press.
Lakoff, G.	1968	Pronouns and Reference, Reprinted in J. D. Mccawley (Ed.) (1976) *Notes From The Linguistic Underground.* New York: Academic Press. pp.275-336.
Langacker, R.W.	1969	On Pronominalization and The Chain of Command, in Reibel, D. A. & Schane, S. A. (Eds), *Modern Studies in English.* pp.160-186. Englewood Cliffs, New Jersey: Prentice-Hall.
Langacker, R.W.	1987	*Foundations of Cognitive Grammar, Vol. 1: Theoretical Prerequisites.* Stanford: Stanford University Press.
Langacker, R.W.	1991	*Foundations of Cognitive Grammar, Vol. 2: Descriptive Application.* Stanford: Stanford University Press.
Lasnik, H.	1976	Remarks on Coreference. *Linguistic Analysis,* 2, pp.1-22
Leech, G.N.	(1991) 1992	Corpora and Theories of Linguistic Performance, in Svartvik, J. (Ed.) *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82nd, Stockholm,* 4-8 August 1991.
Leech, G.N. and Garside, R.	1991	Running A Grammar Factory: The Production of Syntactically Analysed Corpora Or "Treebanks", in Johansson S. and Stenström, A. *English Computer Corpora: Selected Papers and Research Guide.* Berlin: Mouton De Gruyter, 1991.
Leech, G.N. and Short, M.	1981	*Style in Fiction,* Longman, London
Levinson, S.C.	1983	*Pragmatics.* Cambridge University Press.
Lyons, J.	1977	*Semantics. Vol.1 & 2.* Cambridge University Press.
Marx, K.	1976	*Capital: A Critique of Political Economy Vol.1.* Penguin.
McCray, A.	1980	The Semantics of Backward Anaphora, in J. T. Jensen (Ed.) *Cahiers Linguistique d'Ottowa.* pp.329-343.
McKoon, G., Gerrig, R. J., and Greene, S.B.	1996	Pronoun Resolution Without Pronouns: Some Consequences of Memory Based Text Processing, *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Washington, DC 1996 July, 22:4, 919-32
Mitkov, R.	1994	A New Approach For Tracking Center. *Proceedings of The International Conference "New Methods in Language Processing",* UMIST, Manchester, UK
Mitkov, R.	1995	Anaphora Resolution In Natural Language Processing and Machine Translation, *IAI Working Paper.* Saarbrucken
Mittwoch, A.	1983	Backward Anaphora and Discourse Structure. *Journal of Pragmatics* 7: 129-139.
Palmer, F.	1981	*Semantics.* Cambridge University Press.
Popper	1966	*The Open Society and Its Enemies (5th Edn),* Routledge & Kegan Paul.

Prince, E.F.	1981	Toward A Taxonomy of Given-New Information, in (Ed.) Cole. P. *Radical Pragmatics*, New York: Academic Press.
Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J.	1985	*A Comprehensive Grammar of The English Language (CGEL)*, London: Longman.
Radford, A.	1988	*Transformational Grammar: A First Course*, Cambridge University Press.
Reinhart, T.	1976	*The Syntactic Domain of Anaphora. Ph.D. Dissertation*, Cambridge, Mass.: MIT.
Reinhart, T.	1981	Definite NP Anaphora and C-Command Domains. *Linguistic Inquiry* 12:4
Reinhart, T.	1984	*Anaphora and Semantic Interpretation*. London: Croom-Helm.
Ross, J.R.	1967	On The Cyclic Nature of English Pronominalization . in Reibel. D. A. & Schane, S. A. (Eds), *Modern Studies in English: Readings in Transformational Grammar*. pp.187-200. Englewood Cliffs, New Jersey: Prentice-Hall.
Sampson, G.	1987	Probabilistic Model of Analysis. in Garside et al, 1987
Sanford, A.J. and Garrod, S.C.	1989	What, When, and How?: Questions of Immediacy in Anaphoric Reference Resolution. *Language and Cognitive Processes* 1989. 4.(3/4) pp. 235-262.
Searle, J.R.	1979	*Expression and Meaning*. Cambridge University Press.
Segal, E.M.	1995	Narrative Comprehension and The Role of Deictic Shift Theory. In Duchan, J.F. Bruder .G.A. Hewitt, L.E. *Deixis in Narrative* Lawrence Erlbaum Associates New Jersey. 1995.
Short, M.	1996	Exploring The Language of Poems, Plays and Prose. Longman. London
Sidner, C.	1983	Focusing in the Comprehension of Definite Anaphora, Chapter 5 in Brady, M. and Berwick R.C. (eds.), (1983) Computational Models of Discourse, Cambridge, Mass., MIT Press
Sørensen, K	1982	Some Observations on Pronominalization. in *English Studies: A Journal of English Language and Literature*, Lisse. Netherlands (ES). 1981 Apr.. 62:2. 146-155
Stockwell, P.	1995	How to Create Universes With Words: Referentiality and Science Fictionality. *Journal of Literary Semantics*, 23: 159-89.
Strawson, P.F.	1950	On Referring. *Mind* 54
Takami, K.	1985	Definite NP Anaphora and Adverbial Clauses: Argument Against Syntactic Analyses. *Linguistic Analysis*, 15:269-303.

Van Hoek, K.	1997b	*Anaphora and Conceptual Structure*. Chicago: University of Chicago Press.
Van Hoek, K.	1997a	Backwards Anaphora As A Constructional Category, *Functions of Language* 4, 1:47-82
Wales, K.	1996	*Personal Pronouns in Present-Day English*, Cambridge University Press.
Walker, M.A., Joshi, A.K. and Prince, E.F.	1998	Centering in Naturally Occurring Discourse: An Overview, In, Walker, M.A., Joshi, A.K. and Prince, E.F. (Eds.), *Centering Theory in Discourse*, Oxford University Press, 1998.
Webber, B.L.	1981	Discourse Model Synthesis: preliminaries to reference, in Joshi, A.K., Webber, B.L. and Sag, I.A. (eds.) *Elements of Discourse Understanding*, Cambridge University Press. 1981