INFORMATION TO USERS

This reproduction was made from a copy of a manuscript sent to us for publication and microfilming. While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. Pages in any manuscript may have indistinct print. In all cases the best available copy has been filmed.

The following explanation of techniques is provided to help clarify notations which may appear on this reproduction.

1. Manuscripts may not always be complete. When it is not possible to obtain missing pages, a note appears to indicate this.

2. When copyrighted materials are removed from the manuscript, a note appears to indicate this.

3. Oversize materials (maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or in black and white paper format.*

4. Most photographs reproduce acceptably on positive microfilm or microfiche but lack clarity on xerographic copies made from the microfilm. For an additional charge, all photographs are available in black and white standard 35mm slide format.*

*For more information about black and white slides or enlarged paper reproductions, please contact the Dissertations Customer Services Department.
UNIVERSITY OF CALIFORNIA

Los Angeles

Referent Introducing and Tracking in Chinese Narratives

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Linguistics

by

Ping Chen

1986
The dissertation of Ping Chen is approved.

George Bedell

Jack Du Bois

Hung-Hsiang Chou

Sandra A. Thompson, Committee Chair

University of California, Los Angeles

1986
TO THE MEMORY OF MY DEAR FATHER
TABLE OF CONTENTS

Chapter 1  INTRODUCTION..............................................1
  1.1  Goal of the investigation....................................1
  1.2  Brief survey of literature.....................................3
  1.3  Basic assumptions and methodology...........................8
  1.4  Organization of thesis........................................11

Chapter 2  INTRODUCING REFERENTS INTO DISCOURSE.................13
  2.1  Preliminary..................................................13
  2.2  Definition of some basic concepts..............................14
      2.21  Preliminary..............................................14
      2.22  Definite vs. indefinite..................................15
      2.23  Referential vs. nonreferential.........................22
      2.24  Identifiable vs. nonidentifiable......................28
      2.25  Specific vs. nonspecific...............................32
      2.26  Generic vs. individual..................................33
  2.3  Formal encodings of semantic and pragmatic categories.......35
      2.31  Formal encodings of referential vs. nonreferential....35
      2.32  Formal encodings of identifiable vs. nonidentifiable..36
      2.33  Formal encodings of specific vs. nonspecific..........43
      2.34  Formal encodings of generic vs. individual.............43
  2.4  Initial mention of referents in discourse....................44
      2.41  Preliminary..............................................44
Chapter 2

2.42 Indefinite encoding of initial-mention referents... 45
2.43 Definite encoding of initial-mention referents... 49
2.44 Indeterminate lexical encoding... 66
2.5 Indefinite encodings for non-initial mention referents... 71
  2.51 Preliminary... 71
  2.52 Remedial reintroduction... 71
  2.53 Switch of perspectives... 74
2.6 Summary... 76

Chapter 3

STRUCTURE OF NARRATIVE DISCOURSE... 79
3.1 Preliminary... 79
3.2 Taxonomy of discourse... 80
3.3 Structure of narrative discourse... 82
  3.31 Preliminary... 82
  3.32 Cognitive level of narrative discourse... 84
  3.33 Linguistic level of narrative discourse... 86
    3.331 Preliminary... 86
    3.332 Syntactic organization... 86
    3.333 Rhetorical organization... 89
3.4 Summary... 104

Chapter 4

REFERENT TRACKING IN CHINESE NARRATIVE DISCOURSE 105
4.1 Preliminary... 105
4.2 Anaphoric types in Chinese... 106
  4.21 Nominal anaphora... 106
  4.22 Pronominal anaphora... 108
4.67 Summary..............................................226
4.7 Use of nominal anaphora.............................227
  4.71 Preliminary.........................................227
  4.72 NA for maintained and switched reference subject......229
  4.73 NA elsewhere in clause..................................237
  4.74 Explaining the choice of NA.............................241
4.8 Summary..............................................243

Chapter 5 CONCLUSION....................................245

BIBLIOGRAPHY.............................................249
LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL</td>
<td>classifier</td>
</tr>
<tr>
<td>CRS</td>
<td>Currently Relevant State</td>
</tr>
<tr>
<td>CSC</td>
<td>complex stative construction</td>
</tr>
<tr>
<td>DUR</td>
<td>durative aspect</td>
</tr>
<tr>
<td>EXP</td>
<td>experiential aspect</td>
</tr>
<tr>
<td>NOM</td>
<td>nominalizer</td>
</tr>
<tr>
<td>PAR</td>
<td>Particle</td>
</tr>
<tr>
<td>PFV</td>
<td>perfective aspect</td>
</tr>
<tr>
<td>REx</td>
<td>Response to Expectation</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

My deepest thanks go to Sandy A. Thompson, my committee chair, whose intellectual influence on me has been profound and whose help in innumerable ways has been tremendous during the past five years. Her penetrating insights into the linguistic issues have been a constant source of inspirations for me in my own researches. She spent large amount of time reading the several earlier versions of the present thesis, and offered a lot of very helpful comments. Without her help, the thesis won't be as it is.

I am particularly indebted to Jack Du Bois, whose intellectual influence on me is also obvious throughout the present study. I borrowed heavily from his methodology of discourse analysis. Also, he read the draft version of the thesis very carefully, and provided me with many insightful comments.

I would also like to express my sincere thanks to George Bedell, who teaches me how to look at the same phenomenon from different perspectives, and broadens my viewpoint with his erudition.

Thanks are due to the other two members of my committee, Professor Hung-Hsiang Chou and Professor Russell Campbell for being willing to be in my committee and read my dissertation.
I would also like to express my thanks to my fellow students at UCLA, Joelle Bailard, Susanna Cumming, Shoichi Iwasaki, Masao Johnson, Hyosang Lee, Thomas H-T Lee, Feng-hsi Liu, Christian Matthiessen, Galust Mardirussian, Ren Hongmo, Nobuko Sugamoto, and Eric Zee, for their intellectual stimulation and sincere friendship.

I am most grateful to Talmy Givón for his warm help to me in many, many way during the past few years.

I wish to take this opportunity to extend my special thanks to Professor Lu Shuxiang, my mentor, for his unremitting care and invaluable help during the past decade.

I express my heart-felt thanks to my dear parents, who brought me up, and, gave me the best possible education during that unusual period of time.

My thanks go to my wife, Jean, for her constant love and support.
VITA

May 24, 1954
Born, Nanjing, China

1978-1981
M. A., Graduate School of Chinese Academy of Social Sciences, Beijing

1981-1983
M. A., University of California, Los Angeles

1981-1986
Research Assistant, Department of Linguistics University of California, Los Angeles

1983-1986
Teaching Assistant, Department of Linguistics University of California, Los Angeles

1985
Phi Beta Kappa Award

PUBLICATIONS


ABSTRACT OF THE DISSERTATION

Referent Introducing and Tracking in Chinese Narratives

by

Ping Chen

Doctor of Philosophy in Linguistics
University of California, Los Angeles, 1986
Professor Sandra A. Thompson, Chair

The present study investigates how referents are introduced into Chinese narrative discourse and how they are tracked throughout the discourse in terms of various anaphoric devices. Referents on their first mention in discourse may receive one of three lexical encodings in two broad categories: determinate (indefinite or definite), and indeterminate. The encodings are correlated with different assumptions about the identifiability of the referents. Through discourse analysis, it is found that the choice in this respect depends heavily on the saliency of the referents in discourse. After being introduced into the discourse, the referents are subject to one of the three major types of anaphoric encodings, zero anaphora, pronominal anaphora, and nominal anaphora. The study focuses on the discourse-pragmatic features that characterize each of the anaphoric
devices. In a typology of referent tracking mechanisms, little has been asserted for the inference system as exemplified by Chinese. As a case study, the present thesis presents findings that lead to a deeper understanding of this system, and suggests further inquiries along the line. Moreover, the findings reported here show how the important domain of the Chinese grammar must be understood and explained from a discourse-pragmatic perspective.
CHAPTER 1

Introduction

1.1 Goal of the Investigation

The present thesis studies how referents are introduced into Chinese narrative discourse and how they are tracked throughout the discourse in terms of various anaphoric devices.

Referent introduction and referent tracking are two intimately related consecutive phases of the process to verbalize objects (subsumed under the term are human beings, animals, inanimates, etc.) in discourse. When the speaker makes first mention of an object to introduce it into discourse in explicit terms, he is faced with the task of choosing from several options for the encoding of the referent. The same is true with the referent that has been established earlier in the discourse when later mentions are made of it. In such situations, the speaker has at his disposal a repertoire of referent-introducing and referent-tracking devices for his purpose. How the choices are made from the wide range of linguistic devices has attracted the attention of linguists, psychologists, language teachers, researchers in artificial
intelligence, etc., who share an interest in the organization of discourse in the process of language communication.

With my major focus on Chinese linguistics and on general theories of grammar and discourse, I intend to achieve a two-fold goal through the research to be reported in later chapters. On the one hand, the present thesis aims to establish the semantic, syntactic, and discourse-pragmatic properties characteristic of various referent-introducing and referent-tracking, or anaphoric, devices in Chinese, with special attention to the discourse pragmatics-controlled features of the devices. On the other hand, it aims to contribute, in a modest way, to a general understanding of discourse organization in verbal communication through the detailed case study of a set of Chinese narratives. It is my belief that a general linguistic theory, be it in phonetics or phonology, syntax, semantics, or pragmatics, is considered to be a valid one only to the extent that it is supported by evidence from a wide range of languages.

Next, I will discuss briefly the status of the referent-tracking mechanisms of Chinese in the typological treatment offered by Foley and Van Valin (1984).

In the typology of referent-tracking mechanisms, Foley and Van Valin (1984) have identified the following four basic systems from which a given language may resort to one or more for signalling the reference relations of the noun phrase arguments in discourse:
I. Pragmatic pivot in combination with voice opposition

II. Switch reference

III. Gender system

IV. Inference system

The fourth type of reference-maintenance system is claimed to be exemplified by the languages of Southeast and East Asia, such as Chinese, Japanese, Thai, etc., which are characterized primarily by the lack of any of the first three systems and by the heavy use of zero anaphora (ZA).

Foley and Van Valin (1984) concern themselves with the first three types in their discussion, which is in general clearly presented. The fourth type, however, is elevated to the sacred status of "a fine art" which, unlike the other three types, defies a well articulated exposition for its underlying principles.

Obviously, more in the way of investigation is needed beyond the conferring of the status of a fine art. We are eager to know precisely what mechanisms are at work in a language characterized by the inference system in the typological network. It is hoped that the findings to be presented here, together with other studies along similar lines, will contribute towards a deeper understanding of the general issue of how such discourse functions are performed by means of various linguistic devices in human languages.

1.2 Brief Survey of Literature
Previous investigations in this area have been mostly directed to the referent-tracking mechanisms in languages with the first three systems as offered by Foley and Van Valin (1984). Cf. Halliday and Hasan 1976, Grosz 1977, Linde 1979, Reichman 1979, Givón 1983a, Haiman and Munro 1983, McKeown 1984, Fox 1984, inter alia. Not much attention has been paid to referent-introduction (Chafe 1980 being a noteworthy exception), and still less to either topic in the languages characterized by an inference system (cf. Tai 1978, Hinds 1978, 1983, Li and Thompson 1979, 1981, Chen 1984, Li 1985, etc.). Next, I will present a brief review of recent studies on the topic in Chinese.

Li and Thompson (1979, 1981) are among the first few studies that approach the topic of anaphoric choice from a functional-discourse perspective, which is also adopted in the present thesis. Li and Thompson (1979) have examined the use of third-person zero anaphora in Chinese discourse. They have found that one important factor in determining the choice of pronominal anaphora (PA) or ZA is "conjoinability", which refers to the extent to which a clause constitutes a single unit with the preceding clause. A high degree of conjoinability between two clauses tends to favor ZA occurring in the second clause. Furthermore, they succeed in singling out some of the relevant factors that affect conjoinability between clauses.

Li and Thompson (1981) have expanded the scope of enquiries to include PA as well as ZA. They have identified highlighting as the determining factor for anaphoric choice between PA and ZA when the
choice doesn't affect the identification of the referent. They note that it is not enough just to say that whenever the referent can be understood or figured out, ZA can be used, because there are times when it would be quite obvious what the referent for a given pronoun would be if it were omitted, and yet it must be used. When the referent can be understood on the basis of other linguistic and extra-linguistic information provided by the text, a pronoun is used when there is reason to highlight the referent in the context in which it occurs. The investigations presented in Li and Thompson (1979, 1981) have been a source of inspiration and methodological strategies to me in Chen (1984) and in the present work.

Chen (1984) concentrates upon the use of third-person ZA in discourse. I demonstrate how the use of third-person ZA depends upon the addressee's evaluation of the referent of the anaphora for its predictability and negligibility, which are assessed in terms of independent, empirically ascertainable parameters. It is concluded that ZA is most likely to encode the referent standing high in predictability and negligibility.

Li (1985) is the latest treatment of the use of ZA, PA, and NA in Chinese discourse. She has singled out the need on the part of the addressee to make the hierarchical structure of the discourse as the most important factor in the anaphoric choice. She has correlated ZA, PA, and nominal anaphora (NA), respectively, with the formal demarcations of the three levels of discourse units: the clause, the topic chain, and the paragraph.
On the basis of Li and Thompson (1979, 1981), and Chen (1984), the present thesis expands the scope of enquiries to encompass all three major types of anaphoric forms, i.e., ZA, PA, and NA. At the same time, it will investigate the ways in which the referents of anaphora are introduced into discourse. Compared with previous studies, the present thesis will be more comprehensive in the scope of coverage. As will be discussed in the development of the issue, the introduction of referents into discourse and the subsequent tracking of them are considered to be two closed related stages of one process that are most profitably examined side by side. Also, the present study tries to explore beyond the picture of the distribution of various referent-introducing and referent-tracking devices as offered by the data to identify the most fundamental principles that underlie the choices on the part of speakers for specific linguistic devices to meet the discourse requirements. It is hoped that what is presented in the following chapters will account for relevant phenomena that have either escaped attention in the literature, or are still lacking a principled explanation.

Research in related domains in other languages has been very instructive the present thesis. Among the most noteworthy studies conducted recently are Grosz (1977), Clancy (1980), Du Bois (1980), Reichman (1981), Givón (1983), and Fox (1984), from which I have adopted some assumptions and methods that will be discussed in detail later together with my own enquiries. Theoretically and methodologically, I am much indebted to Du Bois (1980), Clancy (1980), and Givón (1983b, 1984) for their studies of related topics in English and other languages, and
to Grimes (1975), Mann and Thompson (1983, 1985), Longacre (1984), and Fox (1984), inter alia, for the framework they have proposed for discourse studies. Comparison will be made between Chinese and other languages as described in the literature when appropriate.
1.3 Basic Assumptions and Methodology

Introduction of participating referents into discourse constitutes one of the first steps for the thematic development of the narrative. Even a cursory look at our data shows unmistakably that not all referents are introduced into discourse in the same way. On initial mention, some referents are presented as nonidentifiable to the addressee, some are presented as if the addressee could identify the referent from its category in spite of the fact that it has never been explicitly mentioned before. The difference in treatment, I believe, results from the difference in the semantic, syntactic, and particularly discourse-pragmatic features characteristic of the referent at issue. Enquiries into the correlation between the encoding and the characteristics of the referent to be introduced will give us insights into the role of these devices in discourse development.

After the introduction, the referents are traced throughout the discourse in terms of various anaphoric devices. Studies on anaphora can be grouped into two categories in the literature, one from the perspective of anaphoric resolution, and the other from the perspective of anaphora production. The first group investigates how, given a particular token of anaphora, the addressee undertakes a process that results in the correct identification of the reference of the anaphora. The second group studies how various factors interact for the addressee in his choice of an anaphoric encoding for a referent. In other words, the former is addressee-oriented and the latter, addressee-oriented.
The study on anaphora in the present thesis basically belongs to the second category. Moreover, as will be demonstrated later in the discussion, it will show that the two perspectives are intimately related in the sense that the addressee must hold some basic assumptions about the addressee's need for anaphora resolution, and the same is true for the addressee in his assumptions about the addressee in anaphora production.

The following basic assumptions are maintained throughout the enquiries to be developed in the present study of the topic.

Firstly, given the nature of the issue, the choice of referent-introducing and referent-tracking devices may display variations among individuals as a result of the difference in their assumptions and goals of the text in the communicative process (cf. Chen 1984). By no means, however, does it mean that referent tracking is of a random nature that defies analysis. The fact that the discourse analysis of the data demonstrates strong preferences, and sometimes exclusive use, of a certain type of device for a specific kind of referents in specific environments indicates that there are valid principles underlying the choices.

Secondly, in choosing a particular form for a newly-introduced referent or a particular anaphoric form for a referent that has been present in the discourse, the addressee is involved with a multiple-fold task. Taking anaphoric choice for an example, in deciding on the
anaphoric form, obviously he must ensure that the addressee will correctly identify the reference of the anaphora on the basis of the linguistic and extralinguistic information available in the discourse. At the same time, he must conform to the general principles underlying the process of discourse production imposed by the communicative roles of the discourse such as coherence (relevance), economy, and so on. The concomitant requirements that must be satisfied in discourse production constitute the basis for the principles underlying the choice under discussion.

Thirdly, the choice made by the addressee is a result of the interaction of the whole range of linguistic and extralinguistic information, syntactic, semantic, discourse, etc., in the possession of the discourse participants. The discourse information is the most important fundamental information that contributes to the whole process. Here, the major task of the linguist is to pin down precisely what kinds of discourse features are characteristic of a specific referent that contribute to the choice of one encoding type over the others by the addressee in the specific situation. At the same time, he must reveal how the various kinds of information interact to result in the final choice.

The basic methodology of discourse analysis is adopted throughout the present study. All the examples in the data come under examination together with the the environment in which they occur. The conclusions are arrived at on the basis of the evidences drawn from discourse

10
analysis. I choose narratives as the discourse data for my present investigation, which includes twenty Chinese pear stories (cf. Chafe (1980)), several stories by the famous Chinese writer Lao She, as well as a few stories published in Chinese newspapers and magazines. The twenty pear stories have been are subject to detailed statistical analysis, while illustrative examples are culled from the other stories when appropriate. It is open to further investigation how the findings reported here differ from those for other discourse types such as expository discourse, or conversation.

1.4 Organization of Thesis

The thesis is composed of five chapters.

The first chapter provides the preliminaries to the investigation to be conducted in the thesis. The second chapter studies how various referents are introduced into discourse. The third chapter presents a picture of how the discourse is organized to serve as the framework within which the anaphoric choice is to be explicated. The fourth chapter discusses the major types of anaphoric encodings in Chinese that are used to track the referents throughout the discourse, and gives an in-depth analysis of the use of each of the three major anaphoric types in discourse to reveal the distinctive features characteristic of these types. The last chapter summarizes what has been established in the present investigation of the subject.
Note to Chapter One

1. Studies on English data do demonstrate a difference among these discourse types (cf. Fox 1984).
2.1 Preliminary

I start with examining how various characters and objects which are involved in the narratives are introduced into discourse before I investigate how they are traced throughout in the form of various types of anaphora. As a case study, I will conduct a thorough analysis of the twenty pear stories. Also, citations will be selected from other sources of data when necessary.

In the first part of the discussion, I am going to explicate some important concepts. Next, I will examine the lexical encodings of the referents upon their first mention in the discourse and reveal how the various encodings are indicative of the assumptions that the speakers make with regard to the semantic, and pragmatic features of the referents in question. In cases where the variations in the form...
encodings of the referents on first mention show that the narrators of the stories don't hold the same assumptions for all new referents alike, I will identify what underlie the different treatments that the speakers accord to these referents. After that, I will talk briefly about whether all referents, after being introduced into the discourse, are encoded in a way that is systematically different from the way in which they appear for the first time earlier in the story, and how such differences, if any, indicate the change in the assumptions that the speakers hold about these referents. In cases that show deviations from the expected treatment, moreover, I will inquire into the factors that are responsible for the phenomenon.

2.2 **Definition of Some Basic Concepts**

2.2.1 **Preliminary**

Before I come to the analysis of the actual data, it is appropriate to define several basic concepts that will be used in this study. They include the following terms which are most commonly represented in pairs:

- **definite vs. indefinite**
- **referential vs. nonreferential**
- **identifiable vs. nonidentifiable**
- **specific vs. nonspecific**
- **individual vs. generic**
The above terms have been widely used in the linguistics literature in such a way that they themselves need explicit definitions as much as the phenomena they are employed to account for. Linguists differ with regard to the denotations and connotations of these terms so that it is necessary for us to set a clear-cut definition for them right at the beginning of the discussion (cf. Chafe 1976, 1979, 1986, Hawkins 1978, Du Bois 1980, Hopper and Thompson 1980, Givón 1980, 1984, inter alia, for discussions of the above-mentioned terms). What is more important, it is necessary to identify the various devices in Chinese which are typically used to convey the information implied by these terms.

Right at the beginning, I would like to draw a distinction between formal classes on the one hand, and semantic and pragmatic properties, on the other hand, as related to the nouns and pronouns (hence subsumed under NPs) in discourse which are often characterized by the above listed terms. Following the practice in Du Bois (1980), I set aside the terms "definite" vs. "indefinite" as labels for formal categories of the lexical encodings for the NPs, and use the other terms to specify various semantic and pragmatic features that characterize these NPs in Chinese discourse. Furthermore, I will seek to establish the correlations between the formal categories and the semantic and pragmatic features of NPs in Chinese.

2.22 **Definite vs. Indefinite**
The lexical encodings for NPs in Chinese fall into three formal categories: definite, indefinite, and indeterminate. The term "explicit" encoding will be used to denote the first two categories, while "unmarked", for the third category. Under the definite group are subsumed personal pronouns, proper names, NPs preceded by the stressed demonstratives zhe ‘this’, and ne ‘that’, or other nominal determiners containing one of the demonstratives, such as zhexie ‘these’, nexie ‘those’, and so on. Possessed NPs also belong to this group. Consider the underlined words in the following examples:

(1) a. You yi ge... chabuduo sishi sui de nanren zai have one CL about forty year DE man DUR
    zhai bala...
    pick pear

b. Ta zhai de hen xinku.
   he pick CSC very tired

a. "There was a man about forty who was picking pears.
   b. He was very tired with the work."
   (S-10)1

(2) a. Ta zhai le san lou guozi....
    he pick PFV three CL fruit

b. You yi ge ren qian zhe yi tiao xiao [niu],
   have one CL person lead DUR one CL small ox

c. E zou guo zhe ge lubian.
   walk past this CL roadside

d. Zhe ge xiao niu kan-zhe zhexie guozi.
   this CL small ox look:at-DUR these fruit

a. "He picked three bushels of fruit....
   b. There was a person passing by the roadside,
   c. (he) was leading a small [ox].
   d. The small ox was looking at these fruit."
   (S-8)
The indefinite category includes NPs with indefinite determiners such as \textit{ji} 'some', \textit{shei} 'someone', \textit{shenme} 'something', and NPs preceded by a few determiners containing \textit{yi} 'one' as a component, such as \textit{yixie} 'some', \textit{yidian} 'a bit (of)', or simply preceded by a classifier. Following is an example:

(3) \begin{center}
\textbf{You ge tizi da zai shu shang.}
\textit{have CL ladder lean at tree up}
\end{center}

"There is a ladder leaning against the tree." (S-6)

The third category includes all those NPs which don't belong to either of the first two categories. In other words, NPs in this category don't receive any explicit lexical encodings, either definite or indefinite as specified above. In this class are mainly bare NPs (including the \textsc{DE(NOM)} construction), NPs used with numerals or unstressed demonstratives \textit{she} or \textit{ne}. Following are some examples:

(4) \begin{center}
a. Ta... gen renjia chajianerguo,
\textit{he with other brush:past}
b. maozi diao xialai.
\textit{hat drop down}
\end{center}

a. "He brushed past the other,
b. and the hat fell down."

(S-9)

(5) \begin{center}
Qianmian ganghao you san ge xiaoheizi.
\textit{front by:chance have three CL kid}
\end{center}

"There happen to be three kids in front."

(S-7)
(6) Ranhou san ge xiao'ai zou guoqu.
    then three CL kid walk past

"Then the three kids walked past."  (S-4)

(7) Ta jiu ba zhai guo de fang xialai.
    he then BA pick EXP NOM put down

"He then put down those picked ones."  (S-12)

The three categories are mutually exclusive although prenominal determiners which mark the NP as a member of one category or another are sometimes found co-occurring with the same NP, as exemplified in the following sentences:

(8) Ta ba ta de ji lou guozi fang zai di shang.
    he BA he DE several CL fruit put on ground up

"He put his several baskets of fruit on the ground."  (S-15)

(9) Ta de liang ge lanzi, zemme shao le yi lan.
    he DE two CL basket how less FFV one CL

"How was it that one of his two baskets disappeared."  (S-9)

For such cases, it is the leftmost determiner of the noun that determines the category type of the NP. Both NPs in the above two sentences, as a result, are considered to be definite due to the presence of the possessive ta de 'his' as the leftmost determiner.

The most striking feature that characterizes these three categories of lexical encodings is the way in which they are correlated with distinctive interpretations as identifiable, or nonidentifiable, with
regard to the identity of the referents as assumed by the speakers for their audience. While referents bearing explicit lexical encodings are assumed to be interpreted as identifiable or nonidentifiable on the part of the addressees, the NPs belonging to the third category depend upon syntactic or discourse contexts for their interpretation in this respect. These generalizations will be treated in detail later in this chapter.

Each of the three formal categories can be broken into several sub-groups. Some sub-groups of the explicit encoding categories are subject to restrictions with regard to the syntactic roles they play in the clause. For illustration, let’s consider the use of the NP with a one-argument dynamic verb such as $lai$ ‘come’, or $qu$ ‘go’. In general, definite NPs occupy the pre-verbal position while indefinite ones occupy the post-verbal position, as exemplified in the following constructed sentences:

(10) \begin{align*}
\text{Ta} & \quad \text{lai le.} \\
\text{he} & \quad \text{come CRS} \\
\text{"He has come."}
\end{align*}

(11) \begin{align*}
\text{Ta gege} & \quad \text{lai le.} \\
\text{he elder:brother} & \quad \text{come CRS} \\
\text{"His elder brother has come."}
\end{align*}

(12) \begin{align*}
\text{Lao Wang} & \quad \text{lai le.} \\
\text{Old Wang} & \quad \text{come CRS} \\
\text{"Lao Wang has come."}
\end{align*}
(13) The san ge ren lai le.
    this three CL person come CRS
    "The three people have come."

(14) Lai le (yi) ge ren.
    come PFV one CL person
    "There comes a guest."

(15) Lai le ge shemme ren.
    come PFV CL some person
    "There comes someone."

For some of the underlined NPs above, reversing the pre- and post-verbal positions will result in unacceptable sentences, as illustrated by the following constructed sentences:

(16) *Lai le ta.
    come PFV he

(17) ??Lai le ta de gege.
    come PFV he DE elder:brother

(18) *Ge keren lai le.
    CL guest come CRS

(19) ??Shemme ren lai le.
    Some person come CRS

On the other hand, however, NPs bearing indeterminate lexical encodings are less subject to such constraints. With verbs such as the one used above, NPs of the third category occur freely in the pre- or post-verbal position in the predicate, depending upon the position for their interpretation as identifiable or nonidentifiable. Consider the following sentences:
(20) San ge xiaochai jixu zou xiaqu.
three CL kid keep:on walk down

"The three kids walked on." (S-6)

(21) Ganghao lai le linwai san ge nanhaizi.
just:now come PFV other three CL boy

"There happened to be three other boys coming." (S-8)

With the information on the formal encodings as background, I am going
to discuss in greater detail the correlations between the formal
categories and the semantic and pragmatic interpretations of NPs in
Chinese discourse.
2.23 **Referential vs. Nonreferential**

Now let's consider the distinction between referential and nonreferential.

In the relevant literature in the domain much attention has focused upon the topic of referentiality both from a linguistic perspective, as well as from a psychological or philosophical perspective. However, much of the discussion found in the literature revolves around single artificial sentences considered out of context, and is only partly related to the goals I set to achieve here. As a discourse analyst, I am much more concerned with reference relations as a mapping from the language to some Universe of Discourse instead of as a mapping of propositions or terms in a language onto the Real World. For the difference between the two, cf. Givón (1984: 388).

Following the discourse-pragmatic tradition as initiated by Chafe (1976, 1979), Hopper and Thompson (1980), Du Bois (1980), Givón (1982, 1984), *inter alia*, referentiality is considered here to be a pragmatic as well as semantic property of mentions of objects in discourse. A linguistic expression is said to be referential when it is used to refer to a particular individual of the description which the speaker assumes may have been established already for its identity or is intended to be established as a new file in the Universe of Discourse.
Taken by themselves, expressions are "referentially opaque" in some contexts, as the underlying phrase in the following sentence:

(1) He is looking for a book on cosmology.

In uttering (1), the speaker may have a specific book in mind, or he may not. For the first case, the phrase is referential, while for the latter case, it is nonreferential.

Situated within discourse, however, the referential opaqueness as involved in (1) considered in isolation evaporates. Du Bois (1980:210) gives two overt means to distinguish non-referential from referential in English discourse.

1. The form of a nonreferential mention is not responsive to presence or absence of a prior mention.

2. The form of a nonreferential mention is not responsive to the semantic distinction between singular and plural.

The underlined phrases in the following sentences are examples of nonreferential use:

(2) Mary's a forester. She's been a forester for three years now.

(3) I finally found out what the best is. I have a Mercedes --- three of them in fact.
Nouns in nonreferential use in Chinese most typically assume lexical encodings other than the definite type. Broadly speaking, they include nouns in the following environments:

1. NPs in attributive use such as in compounds, as categorizing predicatives, in certain comparative constructions, and so on, as illustrated by the underlined words in the following expressions:

(4)  
\[
\begin{array}{ll}
\text{bala-shu} & \text{ji-dan} \\
\text{pear-tree} & \text{chick-egg} \\
\text{yang-mao} & \text{fangyan diaocha} \\
\text{wool} & \text{dialect survey}
\end{array}
\]

(5)  
\[
\text{Lao Zhang de dïdi shi dang-bing-de.}
\]
\[
\text{Old Zhang DE brother be soldier}
\]
\[
\text{"LZ's brother is a soldier".}
\]

(6)  
\[
\text{Ta yang de gou zhan qilai you yi ge ren}
\]
\[
\text{he keep DE dog stand up have one CL person}
\]
\[
\text{neme gao.}
\]
\[
\text{that high}
\]
\[
\text{"His dog stands as high as a human being".}
\]

2. NPs that fall within the scope of negation, as exemplified by the underlined expressions in the following sentences:

(7)  
\[
\text{Wo zhexie tian mei mai shenme shu.}
\]
\[
\text{I these day not buy any book}
\]
\[
\text{"I have not bought any books these days."}
\]
(8)  Ner lian zhang fangzhuo dou meiyou.  
there even CL dining table even not:have.  
"There isn't even a dining table there."

3. NPs in Verb + Object or Preposition + Object conflations. By conflations we mean the syntactic structures in which the NP object is so intimately integrated with the verb or preposition syntactically and semantically that the two are considered to have conflated into a single unit (cf. Hopper and Thompson 1980, and Du Bois 1980). The NPs in the conflations assume the form of simple NP, as in du shu 'read book', chao jia 'have quarrel', da zhang 'fight battle', xi zao 'take bath', and so on. Apparently, such conflations in Chinese don't differ much from the regular verbal or prepositional phrases. Upon close scrutiny, however, some subtle distinctions are still discernable both in syntax and semantics.

Syntactically, the distinction is manifested in several ways. For example, a measure phrase composed of a numeral and a classifier can be inserted into the verb and the noun with the Verb + Noun conflations, but not with regular Verb + Noun combinations.

(9)  du le liang nian shu  
read PFV two year book  
"read for two years"

(10) *du le liang nian she ben shu  
read PFV two year this CL book  
"read this book for two years"
The NP in (10) must be preposed to make the sentence grammatical, resulting in (11):

(11) zhe ben shu du le liang nian
    this CL book read PFV two year
    "read this book for two years"

Following are some more examples:

(12) da le ji tian zhang
    fight PFV some day fight
    "fight for some days"

(13) *da le ji tian na chang zhang
    fight PFV some day that CL fight
    "fight the battle for some days"

Also, the NP in (13) cannot be separated from the verb by a measure phrase. The sentence should be rephrased as (13)’:

(13)’ na chang zhang da le ji tian
    that CL battle fight PFV some day
    "fight the battle for some days"

Semantically, the confluences with the nonreferential NPs are sometimes more idiomatic than the regular Verb + Noun constructions. Consider the following pair of constructed sentences:

(14) Women xia ge yue yao kao yanjusheng.
    we next CL month want exam graduate student
    "We are going to give the graduate students an exam next month".
We are going to sit in the exam for graduate student enrollment next month."

(14) and (15) are homophonous. *Yanjiusheng* 'graduate student' in (14), which stands in a patient-action relationship with the verb, is interpreted in a literal way, while the conflation of the verb and the nonreferential noun in (15) sounds more like an idiomatic phrase in that it deviates from the prototypical relationship between a noun and its object (cf. Givón, 1984). The noun in (14) can be replaced by similar expressions such as *daxuesheng* 'undergraduate', but not in (15). It is in referential use in (14), but in nonreferential use in (15). For a penetrating discussion of similar phenomena in other languages, see Hopper and Thompson (1980).

Note that in isolated sentences when the NP in the same lexical form as those in the *Verb + Noun* appears in the pre-verbal position, it usually acquires the status of being referential and identifiable (to be detailed later) unless it is within the scope of negation or preposed for contrast. Consider the following sentences:

(16) Tamen zhengzai du shu. they DUR read book "They are reading."

(17) Shu tamen zhengzai du. book they DUR read "They are reading the book."
(18) Tamen da wan zhăng le.
they fight end battle CRS

"They have ended fighting."

(19) Zhăng da wan le.
battle fight end CRS

"The battle is ended."

In contrast to the underlined NPs in (16) and (18), shu 'book' and zhăng 'battle' in (17) and (19) are in referential and identifiable (to be detailed later) use.

Chafe (1984) distinguishes three states of activation that a concept may be at a particular time: active, semi-active, or inactive. According to him, concepts as represented by shu 'book' and zhăng 'battle' in (17) and (19) respectively are in a semi-active state either through having been fully active at an earlier point in the discourse, or through having been evoked by a schema.

2.24 Identifiable vs. Nonidentifiable

Given a NP in referential use, we will call it identifiable if the speaker assumes that the hearer can establish a link, by whatever means, between the NP and its unique referent. If the speaker cannot assume that the hearer can establish the link, the referent is considered to be non-identifiable.
It must be pointed out that identifiability doesn't always indicate the ability on the hearer's part to relate the NP to the precise identity of the referent. In everyday usage, partial identifiability is sufficient to justify the assumption of identifiability by the hearer. In this respect, it is appropriate to call attention to the "curiosity principle" that is proposed in Du Bois (1980:233), which reads as follows:

A referent is counted as identifiable if it identifies an object close enough to satisfy the curiosity of the hearer.

This principle explains why in the following sentences the living-room wall may be taken as an identifiable referent whether or not the addressee is able to identify precisely which wall is meant, while the encoding of the wall as nonidentifiable in (2) gives the impression of being unnaturally precise:

(1) The boy scribbled on the living-room wall.
(2) He scribbled on a living-room wall.

As Lyons (1977:189) has put it, the point here is that, once any information at all has been supplied about a referent, it can then be treated by the participants as an individual that is known to them both and identifiable within the universe of discourse by means of a definite referring expression. It is not a necessary condition of successful
reference that the addressees should be able to identify the individual being referred to in any sense of "identification" other than this. However, it must be pointed out that the extent to which this standard of identifiability is relaxed varies with entities marked by different degrees of saliency. It seems that more explicit information is needed for some referents than for some others before the status of 'identifiable' can be assigned. This issue will be taken up later in this chapter.

The means by which identifiability is established in the discourse broadly fall into the following two categories:

1. The referent is explicitly mentioned in the discourse.
2. The referent, although never mentioned before, can be inferred for its identity on the basis of the linguistic or extra-linguistic contexts involved.

Chafe (1984) recognizes the same kinds of mechanisms as above that assign semi-active status to concepts in discourse. Roughly speaking, after a referent has been introduced into the discourse, the addressee can usually assume that if he makes subsequent mentions to the referent, the addressee is able to assign later mentions to their referent, as exemplified in the following sentence:

(3)a. You yi ge ren zai la yi zhi yang.

have one CL person DUR pull one CL goat
b. Ne yang hai bu ken zou.
   that goat still not willing go

c. ㄕ Jiu yongli la 0.
    then hard pull

a. "There was a person pulling a goat.
b. The goat won't go,
c. The (he) pulled (it) hard."

Two referents are introduced in (3)a. The goat is referred to in terms of the definite NP in (3)b. and zero anaphora in (3)c., and the person by zero anaphora in (3)c.

Furthermore, the identifiability of the referent can be acquired through the specific "frame" where it is embedded, or through being anchored to some other definite referents in the discourse. I will discuss this in detail later in connection with the examples from my data.

As can be inferred from the basis on which the addressee assumes identifiability of the referent on the part of the addressee, the concept of identifiability involves a continuum which ranges over varying degrees of the feature, which are in turn correlated with various encoding devices in the language. This will be my concern in later chapters.
2.25 **Specific vs. Nonspecific**

Nonidentifiable NPs may be in specific use or nonspecific use. The NP is taken to be in specific use when the speaker uses it to refer to a particular, although not necessarily identified, individual in his mind. If he does not have such particular individual in mind when he uses the NP, it is considered to be in nonspecific use. It is open to question whether the NP in nonspecific use is a referring expression or not (cf. Lyons 1977:189). Here again, I follow the practice adopted in Du Bois (1980:224), and recognize a nonspecific, non-identifiable referent in the underlined words *a pear or two* in (1), in contrast to the specific, non-identifiable referents *a kid*, and *a bicycle* in (2):

(1) And you think "wow", this little boy's probably going to take *a pear or two*.

(2) ... *a kid* comes by on *a bicycle*.

The specific vs. nonspecific contrast is not applicable to NPs in nonreferential or generic use. It has been mentioned above that the NPs bearing indefinite encodings in Chinese are always to be interpreted as nonidentifiable, while those in indeterminate encodings depend upon the context for their interpretation in this respect. When the NPs belonging to some sub-groups of the indefinite encodings do occur in the syntactic slots which most typically presuppose identifiability of the referents, such as in BA constructions or in pre-verbal positions, they are to be
interpreted as referring to particular individuals, although still nonidentifiable in reference. Consider the following examples:

(3) Ne ge fangjian feichang leng, ba
    that CL room very cold BA

    ge laotaitai dong de fadou.
    CL old:lady freeze CSC shiver

    "It was so cold in that room that there was an old lady who was shivering with cold."  (People's Daily-1984)

(4) Yi ge xiaohai guolai le.
    one CL kid come CRS

    "There was a kid coming along."  (S-15)

The speakers presuppose the existence of a particular old lady and a kid in uttering (3) and (4), although they cannot identify them for the hearers. These examples suggest that it is more adequate to characterize the NPs in these positions as specific referential, instead of as identifiable, since they must always be specific referential although they can sometimes be nonidentifiable.

2.26 Generic vs. Individual

Generic reference points to a concept in the mind of the interlocutors which is representative of the whole class while individual reference refers to a member or a set of members of the class. With generic reference, as distinguished from individual
reference, the referential vs. nonreferential and identifiable vs. nonidentifiable distinction is neutralized. In a sense, the concept to which a generic reference is made is always there in the mind of the interlocutors. A natural consequence of this is that its use is free from constraints usually imposed upon individuals, such as definite encodings for identifiabes, and indefinite encodings for nonidentifiables, and so on. In Chinese, as in English, NPs bearing any of the encoding types can be used for generic reference. The most commonly used type, however, is the indeterminate encoding. Also, previous mention of the concept doesn't guarantee definite encoding for the later mentions. Consider the following sentence:

(1) a. Fujin hai keyi tingdao yang jiao de shengyin.
around still can hear goat bleat sound

b. Zou yi ge ren, have one person

c. uh, qian she yi tou yang zou guoqu.
uh lead DUR one CL goat walk past

a. "The goat bleats can be heard in the vicinity. ...
b. There was a person
c. uh, (who) walked past, leading a goat."
(S-4)

The goat in (1)b. is presented as brand new without any influence from the previous appearance of the goat in generic use in the preceding clause.
2.3 Formal Encodings of Semantic and Pragmatic Categories

2.3.1 Formal Encodings of Referential vs. Non-referential

I will begin by introducing a new term nondefinite, which is used here to include both indefinite and indeterminate encodings of the NPs.

The NPs encoded in definite terms in Chinese are all referential. Those in nondefinite encodings can be referential or non-referential, depending upon the context. When they appear in syntactic slots such as in BA constructions, or preverbal positions, they are interpreted as referential. On the other hand, when they are in typically nonreferential environments, as discussed in §2.23, they are nonreferential.

However, in positions where the NPs in nondefinite encodings may have two interpretations, such as post-verbal slots, the indefinite encoding is usually adopted if the addressee intends to signal it as referential. Let's consider an example illustrating this point:
A. - a. "The *baihe* will be used as medicine to cure a *kid* of his cough.
   b. This is the popular remedy in our Lake Tai area."

B. - c. "Is he your *kid"?"

---

(People's Daily-84)

In a situation where the *kid* has never been mentioned before, the bare form *xiaohai*, instead of the indefinite form, would be more appropriate in (1)a. if Speaker A doesn't intend to establish a new file in the context shared by the interlocutors. Later, the NP proves to be in high plot saliency (what this expression means will be explicated later), which explains why the indefinite instead of indeterminate encoding is adopted here.

2.32 Formal Encodings of Identifiable vs. nonidentifiable
As discussed in S2.22, the NPs which bear definite lexical encodings are to be interpreted as identifiable, and the NPs in indefinite lexical encodings are to be interpreted as nonidentifiable, no matter what syntactic positions they may assume in the sentence; moreover, some syntactic positions prohibit some specific sub-groups of NPs from bearing explicit lexical encodings. Consider some more examples:

(1)  Tamen zuotian zou le.
     they yesterday go CRS
     "They left yesterday."

(2)  *Zuotian zou le  tamen.
     yesterday go  PFV they

(3)  Wo deng le  zhexie  ren  san ge  xiaoshi.
     I wait  PFV  these  person  three CL  hour
     "I waited for these people for three hours."

(4)  *Wo deng le  shenme  ren  san ge  xiaoshi.
     I wait  PFV  some  person  three CL  hour
     "I waited for some people for three hours."

Although there may be some variations with regard to the acceptability of sentences with the NPs in various definite or indefinite lexical encodings in these positions, one thing that is beyond any doubt is that the NPs in explicit encodings preserves the status of identifiability or nonidentifiability wherever they turn up in the sentence.
The situation is much more complicated with the NPs in indeterminate encodings, which depend upon the context for their interpretation as identifiable or nonidentifiable. First, let’s consider the bare NPs.

It was determined more than a century ago that the same NP is interpreted as identifiable in (5), and nonidentifiable in (6) (cf. Chao 1968, Teng 1975, Zhu 1982, etc.):

(5) \begin{tabular}{llll}
Keren & lai le. \\
guest & come CRS \\
"The guest has arrived."
\end{tabular}

(6) \begin{tabular}{llll}
Lai & keren le. \\
come & guest CRS \\
"There comes a guest."
\end{tabular}

In other words, the addressee has to assume that the addressee knows the identity of the guest(s) when he utters (5), while such presuppositions don’t exist with (6).

Judgements of acceptability of the relevant expressions in the data have led me to the conclusion that the bare NPs also demonstrate strong tendencies for either identifiable or nonidentifiable interpretation when they turn up in some other syntactic positions in the sentence. Following are my observations in this respect.
Bare NPs in BA constructions are assumed to be identifiable in reference, except for the V + NP conflations in which the NP serves as the filler in the slot. Consider the following example:

(7) Ta yijing ba wenzhang xie hao le.
    he already BA paper write well PFV

"He has already finished the paper."

The same is generally true with the bare NPs in the V __ NP slot, although not to as great an extent as in BA constructions. Consider the following example:

(8) Wo song le tongxue ji ben shu.
    I give PFV classmate several CL book

"I gave my classmates some books."

Similarly, a bare NP is usually assumed to be identifiable in reference when it comes in front of another NP as a restrictive modifier. Consider the following example:

(9) Kerewmen de yifu fang zai waimian.
    guests DE clothes put at outside

"The clothes of the guests are put outside."

Bare NPs are to be interpreted as nonidentifiable when occurring after you "exist", or in the V NP __ slot with a ditransitive verb. Consider the following example:
(10) You shan, you lu, hai you shumu.
    have mountain have road still have tree

    "There are mountains, roads, and trees." (S-4)

Basically speaking, NPs preceded by numerals demonstrate the same regularities with regard to the identifiable vs. nonidentifiable contrast when they assume the syntactic positions discussed above, although the correlation is less rigorous and the hearer is left with more leeway to interpret them in either way. Following are some of the examples:

(11) San ge xiahai jixu zou xiaqu.
    three CL kid continue walk down

    "The three kids walked on." (S-10)

(12) Qiammian ganghao you san ge xisananhai.
    front just have three CL kid

    "There happened to be three kids in front." (S-19)

(13) ne ge nanhai, ... san ge nanhai zi
    that CL boy ... three CL boy

    yi ge nanhai zi
    one CL boy

    "that boy,... one of the three boys" (S-3)

(14) Ne xiaohair gei ta san ge ... san ge shuiguo.
    that kid give he three CL three CL fruit

    "The kid gave him three, ... three fruits." (S-4)

The underlined NP in (13), which occurs in the pre-nominal restrictive construction, tends to be interpreted as identifiable, while the
underlined NPs after you in (12), and in V NP ___ slot in (14) are most likely assumed by the speakers to be nonidentifiable in reference. On the other hand, the preverbal NP in (11) is open to interpretation in either way, though most speakers probably assume it to be identifiable for the hearer when it is used in this way.

In the examples given above, the NPs in indeterminate lexical encodings demonstrate strong correlations between the syntactic positions and identifiable vs. nonidentifiable interpretations as I have just discussed; on the other hand, however, there are also some syntactic slots which don't display any significant tendencies in this respect, such as the postverbal position with two-argument verbs, and the pivotal object position. Interlocutors may interprete the equivocal indeterminate NPs in these slots as either identifiable or nonidentifiable. Following are some examples:

(15) Ta you kando san ge nanhaizi keng zhe 
     he again see three CL boy munch DUR
     bala zheyang zou guoqu. 
     pear this:way walk past

     "Then he saw (the) three boys walk past, munching pears."
     (S-2)

(16) Ta turan kando san ge xiaohaizi you ne bian 
     he suddenly see three CL kid from that side
     shumu zou guolai. 
     tree walk over

     "Suddenly, he saw (the) three boys walk over from the trees over there".
     (S-19)
Although the kids in the above sentences have already been introduced into the discourse earlier in the stories, they can be presented here as identifiable or nonidentifiable depending upon which "world" the narrator chooses to adopt (I will come to this point soon). As it is, neither the syntactic positions nor the lexical encodings of the NPs in the above sentences have offered much to suggest either an identifiable or a nonidentifiable interpretation.
2.33 Formal Encodings of Specific vs. Nonspecific

All definite NPs are specific in reference. Furthermore, the specific vs. nonspecific distinction is neutralized with nonreferentials. As a result, what is involved here with the specific vs. nonspecific dichotomy are NPs in nondefinite encodings which are in referential use. In syntactic positions other than those which typically have specific referential NPs, both specific and nonspecific interpretations are possible for referents in nondefinite encodings. Consider the following example:

(1) Ta yao mai (yi jian) xin yifu.
    he want buy one CL new clothes

"He wants to buy a new outfit."

(Yi jian) xin yifu can be taken either as referring to a particular piece of new clothes, or as referring to any new clothes.

2.34 Formal Encodings of Generic vs. Individual

There is no significant correlation between individual reference and any one type of formal encodings in Chinese. As can be inferred from the above discussions, various encodings can be used for individual reference. For generic reference, on the other hand, the bare NP is by far the most commonly used encoding form. In the Chinese
pear stories, it is found to be the sole encoding type for objects in
generic use.

2.4 Initial Mention of Referents in Discourse

2.4.1 Preliminary

So far I have explicated the relevant semantic and pragmatic
features of NPs in discourse, and discussed how these features are
correlated with morpho-syntactic encodings in Chinese. Next I will
examine the data to see how the referents are introduced into discourse
in the twenty pear stories.

The examination of the data shows that the referents make their
debut in the discourse in several different ways. I will divide the
initial-mention referents into three groups according to whether they
bear determinate (definite or indefinite), or indeterminate encodings,
and subject all the tokens in the three groups to meticulous scrutiny in
order to identify the morphological, syntactic, and semantic features
characteristic of the referents in each group. It is hoped that the
results obtained here will shed some light upon the cognitive contraints
which are at work when speakers make their choice as to how to introduce
the referents into the discourse.

44
2.42 **Indefinite Encoding of Initial-Mention Referents**

As the first step, I will discuss the initial-mention referents that are encoded as indefinite NPs. 177 instances of objects or individuals are found to be introduced in this way. All of them are referential nonidentifiables. They fall into three sub-groups according to their lexical features:

<table>
<thead>
<tr>
<th>Sub-group</th>
<th>Lexical Feature</th>
<th>Number of Token</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>yi (CL) 'one', yidian 'a bit', yixie 'some'</td>
<td>+ NP 172 (97%)</td>
</tr>
<tr>
<td>2</td>
<td>ji 'several'</td>
<td>+ (CL) + NP 2 (1%)</td>
</tr>
<tr>
<td>3</td>
<td>shei 'somebody', shenme 'something'</td>
<td>3 (2%)</td>
</tr>
</tbody>
</table>

**TABLE 2:1**

Subgroupings of Indefinite NPs Encoding Initial-Mention Referents

As shown in Table 2:1, the great majority of the referentials lexically encoded as nonidentifiable in the Chinese peer stories appear together with the numeral "one". Considering what is involved in introducing a new referent into discourse, this is no coincidence. According to Givon (1984), the speakers are confronted with a complex task when they usher in a new referent. They both have to imbue the new topic with referentiality, and at the same time have to identify it as one member
within the type. "One" is a most appropriate candidate for this double function.

The distribution of these indefinite referents among the syntactic slots is displayed in Table 2:2:

<table>
<thead>
<tr>
<th></th>
<th>V</th>
<th>BA</th>
<th>V</th>
<th>shi(V)</th>
<th>V</th>
<th>V</th>
<th>NP</th>
<th>Prep</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub-group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>64</td>
<td>9</td>
<td>2</td>
<td>82</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>97%</td>
<td>100%</td>
<td>100%</td>
<td>97%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td></td>
<td>100%</td>
</tr>
<tr>
<td>(2)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>66</td>
<td>9</td>
<td>2</td>
<td>84</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 2:2**

Syntactic Distribution of Initial Mention Indefinites

You, which literally means 'have' in Chinese, is the word that is commonly used to present a nonidentifiable referent in Chinese discourse. As we can see from Table 2:2, the position after you and the post-verbal slot are the two most common syntactic positions for newly introduced referents to occupy. Similar situations have been reported in other languages as well (cf. Clancy 1980, Du Bois 1980, Givón 1979a, inter alia). Following are some examples:
(1) a. You yi ge nongren,  
    have one CL farmer  

    b. ō zai lu pang de yi ke shu shang zhai guozi.  
    in road side NOM one CL tree up pick fruit  

    a. "There was a farmer,  
    b. (who) was picking fruit up in a tree by the roadside."  
    (S-5)  

(2) a. Cong nebian guolai yi ge xiaohaizi,  
    from there come one CL kid  

    b. ō qi zhe yi liang zixingche.  
    ride DUR one CL bike  

    a. "From there came a kid,  
    b. (he) was riding a bike."  
    (S-4)  

The other slots see many fewer newly-introduced indefinite NPs. Contrary to the claim that the preverbal position and BA construction don't admit indefinite NPs (cf. Wang 1955, Teng 1975, Zhu 1980, etc.), it is interesting to note that we do find instances in which the NP in indefinite encoding appears in the BA construction \((n = 2)\), and in the pre-verbal position \((n = 9)\).

(3) (Ne ge xiaohaizi)... ba yi lou shuiguuo tongtong  
    that CL kid BA one CL fruit all  

    ban dao che shang.  
    move to bike up  

    "(That kid)... moved the whole basket of fruit to his bike."
    (S-20)  

(4) Yi ge gongren zai ner chai shuiguuo.  
    one CL worker at there pick fruit  

    "A worker was picking up fruit there."  
    (S-3)
The syntactic positions occupied by the underlined words in (3) and (4) typically have NPs bearing definite encodings, or lexically indeterminate NPs which are interpreted as identifiable. The fact that very few indefinites are found in these positions is a good indication that they are not the usual places for nonidentifiables.

The same is true for other syntactic slots which, as I remarked in § 2.22, tend to impose certain restrictions upon the definite vs. indefinite status of of NPs in the positions. Now, note that in Table 2:2 we have only one case in which an indefinite NP is used as a pronominal restrictive modifier, which is presented as following:

(5) (Zhe)... bu xiang zai yi ge jinbu guojia de
    this not like at one CL developed country DE
    yi ge difang.
    one CL place

"(This)... is not like somewhere in a developed country."

(5-4)

However, this sentence sounds very unnatural to me. Taken as a whole, the story narrated by the fourth speaker impresses me as very stilted and artificial. In connection with the fact that indeterminate NPs in these positions tend to be interpreted as identifiable, the claim is justified that some syntactic positions in Chinese don’t favor referents in nonidentifiable use.

In this section, I have examined the morphological and syntactic features of all the initial-mention referents which are encoded as
indefinite NPs in the pear stories. The semantic characteristics of the indefinite NPs under survey here will be discussed in connection with the definite and indeterminate NPs in S2.44.

Next, let's consider the initial-mention referents which are encoded in definite terms in the pear stories.

2.43 Definite Encoding of Initial-mention Referents

Against the expectation that all newly introduced referents will be encoded as nonidentifiable, I find quite a few referents in the data which bear definite lexical encodings on their first mention in the discourse. They are presented as identifiable individuals to the hearer although no coreferent NPs have been established earlier in the file to which these NPs can be directly connected for their identities. Table 2:3 gives the lexical subgroupings of these initial-mention definite referents:
<table>
<thead>
<tr>
<th>Sub-group</th>
<th>Lexical Feature</th>
<th>Number of Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ta &quot;he/she&quot;</td>
<td>3 (5%)</td>
</tr>
<tr>
<td>2</td>
<td>she &quot;this&quot; } ( + NP )</td>
<td>14 (25%)</td>
</tr>
<tr>
<td>3</td>
<td>ne &quot;that&quot; }</td>
<td>25 (45%)</td>
</tr>
<tr>
<td>4</td>
<td>restrictive modifier + NP</td>
<td>14 (25%)</td>
</tr>
<tr>
<td></td>
<td><strong>Total:</strong></td>
<td><strong>56 (100%)</strong></td>
</tr>
</tbody>
</table>

**TABLE 2:3**

Lexical Subgroupings of Definite NPs on Initial Mention

Table 2:4 shows how these NPs are distributed in various syntactic slots.
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The question is how it occurs that the referents that have never been introduced into the discourse are presented by the speakers as if they are identifiable for the hearer on initial mention. Examination of all the instances of the identifiable initial mentions has identified three major factors which are responsible for the definite encodings of the initial-mention referents.
The first and the most important factor involves extralinguistic knowledge such that referents may be identifiable even though they are being mentioned explicitly for the first time. Knowledge of this kind includes socially and culturally shared knowledge maintained by the interlocutors involved. The names for heavenly bodies, proper names, and so on, belong to this category. The identities of these objects are stored in the minds of people as if in a permanent file and will be triggered at the mere mention, as exemplified by 要 be ge jinse yangguang 'the golden sunshine' in the following sentence:

(3) Shuye zai 要 be ge jinse yangguang zhi 要 be xia tree:leave in this CL golden sunshine of down shandong. glisten

"The tree leaves were glistening under the golden sunshine."
(S-4)

Extralinguistic knowledge may also derive from the deictic context of the communication which assigns a unique reference to the terms that are related to the discourse through temporal, spacial, or interpersonal relationships involved. Consider the following sentence:

(4) Tamen... dagai shi ne ge difang de haizi. they perhaps be that CL place NOM kids

"They are perhaps the kids living in that area."
(S-7)

(5) Ne ge jinse hen mei. that CL scenery very beautiful

"The scenery was very beautiful."
(S-8)
That place and that scenery acquire the identity from the deictic context of the narration, and unmistakably refer to the place where the story took place and the scenery thereof. Kinship terms, etc., belong to the same category.

Deictic identifiability extends to referents which, instead of acquiring identifiability directly from the context, derive it from other referents which have already been established in the discourse. It is appropriate to mention the "frame" concept here (Minsky 1975). (Cf. "script", "scenario", "schema", etc. as developed in such works as Schank 1972, Fillmore 1975, Minsky 1975, Du Bois 1980, Sanford and Garrod 1981, Chafe 1984. The difference among the terms is not my concern.) It has been proposed that human knowledge is stored in memory in the form of data structures, which are called "frames" by some scholars as Fillmore (1975), Minsky (1975), *inter alia*. Frames are considered to be representative of stereotyped situations. If a frame has been evoked in the discourse, the referents which are typically involved in the situation will be taken as of identifiable reference even if it has never been mentioned before. For instance, the waiter in the frame of a restaurant, or the frying-pan in the frame of a kitchen can derive their definite encodings in this way. Some of the definite initial mentions in our data obviously belong to this category. Consider the following sentence:
(6) a. Ta ... qi che de shihou,
    he ride bike DE time

b. ta zuo shou hai yao fu zhe ne ge bashou,
    he left hand still want hold DUR that CL handle

a. "When he ... was on the bike,
b. he had to hold the handle with his left hand."
(S-6)

Here the handle is unmistakably assumed to be the handle of the bicycle
that the boy is riding. It becomes identifiable through inference once
the frame of a bicycle is evoked in the discourse.

Considering the mechanisms underlying the acquisition of
identifiability by referents in frames, the fact is easily understood
that a majority of the definite initial mentions assume the form of the
NP preceded by the possessive, which serves as the anchor to which the
NP is associated for its identity. Following are some illustrative
examples:

(7) Ta faxian ne xiaohai de maozi diao zai ner.
    he find that kid NOM hat drop in there

"He found the kid's hat down there." (S-7)
(8) a. Ta jiu na zou le yi kuang,
    he then take away PFV one basket
b. shang, put to be NOM bike
c. dai zou le
carry away CRS

a. "He then took one basket,
b. put it on his bike,
c. and carried it away."

(9) a. You yi ge ren dao...
    have one CL person to
b. shang, put to be NOM bike
c. "There was a person...
b. (who) straight went to his orchard."

(S-14)

(S-19)
Second, some referents are encoded in definite terms by a mistake on the part of the narrator (n = 6). Out of carelessness, the referent is presented as identifiable to the listeners when the latter have no means whatsoever to effect a connection between the referent and something already established in the discourse. All three instances of pronominal references belong to this group, and most of the NPs used together with demonstratives are found to be in this group, too. Consider the following example:

(10) Ta gankuai lai bang ta.
    he hurry come help he

"He (one of the threesome) hurried to help him."

(5-17)

Here it is totally impossible for the listener to establish the identity of the subject of the sentence since no introduction has been made of either the paddle boy or the threesome earlier in the story. Most speakers in similar cases came to realize the blunder soon afterwards, and remedied the situation through reintroducing the referent into the discourse. Consider the following example:
(11) a. Suoyi, **ne ge nanhai, san ge nanhaizi limian de**
therefore that CL boy three CL bo\*y Inside NOM

yi ge nanhaizi cai pao huilai,
one CL boy just run back

b. 介 huan gei ta ne ge maozi.
return give he that CL cap

a. "Therefore, that boy, one of the three boys, ran back,
b. and gave the hat back to him."
(5-3)

The presupposition of the identifiability of **ne ge nanhai** 'that boy' is
in no way justified in the text. To remedy the situation, the speaker
immediately reintroduces the referent with a classifying modifier "from
the three boys" which serves as the frame of reference that directs the
hearer to the intended referent.
The third factor accounting for the definite initial mentions involves a syntactic hole in the sentence which presumes the existence of a referent that the speaker was not able to articulate, perhaps due to the low codability of the referent concerned (cf. Downing 1980) (n = 3). Usually we have a deictic pronoun, zhe, or ne, that follows immediately, referring back to the unexpressed referent in the syntactic hole. Consider the following example:

(12)a. Qita ne ji ge xiaoaizi a, ... qizhong you
    other that several CL kid MOD within have
    yi ge, ganghao shi zai wan ..., one CL just be DUR play

b. ne ge jiao shenme?
   that CL call what

   a. "The other several kids, ... one of them was playing with a...
      b. what's that called?"

   (S-15)

The missing argument of the verb wan 'play' leaves a syntactic hole in the sentence from which the deictic ne 'that' derives the identifiability of its reference. It is understood by the interlocutors that that refers to what the boy was playing with.

Of the three factors mentioned above which underlie definite encodings of the initial-mention referents, the extralinguistic knowledge factor is obviously the one that is responsible for the majority of the instances under consideration which are presented as identifiable although they have not been mentioned in explicit terms earlier in the discourse. To borrow a term from Chafe (1984), these
initial-mention NPs presented as identifiable are subsumed under one of the two kinds of semi-active concepts which belong to the set of expectations associated with a schema or frame. When a schema or a frame has been evoked in a narrative, some if not all of the expectations of which it is constituted presumably enter the semi-active state. It is both legitimate and natural to treat these objects in semi-active state as identifiable within the specific schema or frame of the discourse.

So far, we have analyzed the factors which are responsible for the definite encodings of the initial-mention referents. However, it is indispensable to the full understanding of the morpho-syntactic encodings of initial-mention referents to investigate whether there are systematic differences between the initial-mention indefinite NPs and the initial-mention definite NPs, which are assumed to be identifiable on the basis of the extralinguistic knowledge shared by the interlocutors in spite of the fact that it is their first explicit mention in the discourse. Specifically, I want to investigate whether the two groups of NPs are each correlated with a different set of characteristic features, and if so, what underlies the correlation.

First, I will compare the semantic classes found in the two groups of NPs. Table 2.5 displays how the NPs of various semantic groups are distributed among the two types of formal encodings (only groups with more than ten token NPs are presented):

60

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>Definite Encoding</th>
<th>0 (%)</th>
<th>0 (%)</th>
<th>8 (23%)</th>
<th>5 (14%)</th>
<th>19 (79%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indefinite Encoding</td>
<td>95 (100%)</td>
<td>12 (100%)</td>
<td>27 (77%)</td>
<td>30 (86%)</td>
<td>5 (21%)</td>
</tr>
</tbody>
</table>

story character non-human animal moveable inanimate object vegetation clothes

Total: 95 12 35 35 24

Table 2:5
Distribution of Initial Mention NPs Belonging to Various Semantic Classes Between Definite and Indefinite Encodings

The NPs which receive definite encodings due to the second and third factor as discussed above are not included in the counts presented in Table 2:5. All the referents in definite encoding here are assigned identifiability through the frame as earlier explicated. What is evident from the table is that the semantic category of the referent is closely related with whether it is assigned identifiability or not on the basis of the extralinguistic context. Here it is necessary to note the difference in saliency registered by referents in various semantic categories as shown in Table 2:5. Saliency relates to the ways in which the objects present in a situation are seized on by the participants of discourse as foci of attention. Two kinds of saliency are recognized in the linguistic literature (cf. Du Bois, 1980), inherent saliency and plot saliency. Inherent saliency is determined by the intrinsic properties of the referents. For illustration, human and animates tend
to command more attention in discourse than inanimates. On the other hand, plot saliency depends upon the importance of the role played by the referent in the plot development of the discourse. Generally speaking, the NPs which are low in inherent saliency are more likely to acquire identifiability from the extralinguistic context than those high in inherent saliency. This is a common sense observation since it is more likely that the former will be incorporated into a frame than the latter. As the findings reported in Du Bois (1980) show, while the frame of a human being triggers definite initial narrative mode mention of the person’s body parts, the reverse does not hold. In the Chinese pear stories, there doesn’t appear to be a frame that evokes definite mentions of human, or nonhuman animate referents. All of them are presented as nonidentifiable on initial mention, except those for which the second and the third reason as discussed above are responsible.

The NPs of the other three semantic classes, i.e., movable inanimate, vegetation, and clothes, show varying degrees of alternation between the two types of explicit lexical encodings. Of the initial-mention referents, 23% (n = 5) in the semantic category of movable inanimates, 14% (n = 5) in the category of vegetation, and 79% (n = 19) in the category of clothes are encoded in definite terms. For these NPs, it doesn’t mean that there are no regularities in the choices on the part of the speakers. A careful examination of the data demonstrates that there are correlations between the encoding types and other parameters in addition to the semantic classes. One of these parameters is the discourse mode? Of the 24 mentions of NPs referring
to clothes which are in explicit encodings in the data, 19 receive
definite encodings, and only 5 are in indefinite encodings. It is
noteworthy that the 19 definite NPs are all found in narrative mode,
while four out of the five indefinite NPs are in the descriptive mode,
as exemplified by the following two sentences:

(13) Ta dai zhe yi ge maozi.
    he wear DUR one CL hat

"He had a hat on."
(S-4)

(14) (Ta) chuan zhe yi ge bu hen teshu de yifu.
    he wear DUR one CL not very special NOM clothes

"(He) was wearing an outfit which didn’t look very special."
(S-4)

The observation here is in conformity with the report made in Du Bois
(1980) with regard to the English data. It is proposed in Du Bois (1980)
that timing of introduction is involved in the encoding of the NPs. He
suggests that there is a critical introduction period for referents.
When the speaker tries to present a scene to the hearer, he chooses to
introduce the salient object of the scene in the descriptive mode so
that the hearer reconstructs a complete picture of the scene in the
mind. The time when he focuses his attention to this purpose is called
the critical introduction period (Du Bois 1980:249):

If an object has high intrinsic saliency, it is
likely to be mentioned when it first appears,
during its critical introduction period. If it has
high plot saliency, it is likely to be first mentioned wherever it fits into the plot. If this happens to come early, during its critical introduction period, it will be indefinite; if not, it will usually be definite.

As we can see, the above observation also applies to the Chinese data. The hat of the bike boy is an instance of low intrinsic saliency but high plot saliency. When the frame trigger, the bike boy, first appears in the discourse, only Speaker 4 mentions the hat he has on the head. The other speakers simply ignore it during that period. However, when the attention focuses upon the hat later in the story as one of the threesome picked up the hat, returned it to the bike boy, and got three pears for reward, it was encoded as definite by seventeen out of the nineteen speakers who mentioned it in the story.

The relationship between the discourse modes and the critical introduction period is also quite obvious. The mention of the referents in descriptive mode is usually limited within the critical introduction period. Beyond the period, the speakers are much more inclined to rapidly advance the story plot with sentences in narrative mode instead of returning to the description of something introduced earlier in the story.
By now, we have discussed the morphological, syntactic, and semantic features of two groups of initial-mention referents, one encoded in indefinite terms, and the other in definite terms. Next, we turn to the third group of initial-mention referents — those encoded in indeterminate terms.
2.44 Indeterminate Lexical Encoding

In the data there are also quite a few initial-mention referents which don't bear any explicit lexical encodings, neither in definite nor in indefinite terms. As mentioned earlier, these NPs belong to the category of indeterminate lexical encodings, which again fall into several sub-groups, as specified above. In the twenty pear stories, I find 136 tokens of initial mentions lexically encoded as indeterminate. They are distributed between the three sub-groups of indeterminate encoding, i.e., bare NPs, NPs used with numerals, and NPs used with unstressed determiners zhe 'this' and ne 'that' (cf. § 2.22) in the following way:

<table>
<thead>
<tr>
<th>Sub-group</th>
<th>Lexical Feature</th>
<th>Number of Token</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bare NP</td>
<td>84 (62%)</td>
</tr>
<tr>
<td>2</td>
<td>Numeral + NP</td>
<td>41 (30%)</td>
</tr>
<tr>
<td>3</td>
<td>zhe/ne + NP</td>
<td>11 (8%)</td>
</tr>
</tbody>
</table>

Total: 136

TABLE 2:6
Subgrouping of Lexically Indeterminate Initial Mention NPs

The fact that these NPs are not lexically marked for identifiability or nonidentifiability doesn't mean that they are all obscure in this respect. It is true that there are cases where this information is not
clear. But, as mentioned earlier, for most of the NPs bearing this kind of lexical encoding, the syntactic positions that they occupy can be taken as a cue to the assumptions that the speaker holds about the identifiability of the referents for the listeners, although the correlation between the syntactic slots and the identifiable vs. nonidentifiable assumptions is usually not very rigid.

Table 2:7 shows the syntactic distribution of three sub-groups of the initial-mention referents in the pear stories that are encoded in indeterminate terms:

<table>
<thead>
<tr>
<th>Number of Token</th>
<th>you</th>
<th>_V</th>
<th>BA</th>
<th>V_</th>
<th>V NP</th>
<th>V V</th>
<th>Prep</th>
<th>_N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical Subgroup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bare NP</td>
<td>5</td>
<td>8</td>
<td>2</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Numeral + NP</td>
<td>11</td>
<td>4</td>
<td>0</td>
<td>12</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>zhe/ne + NP</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total:</td>
<td>16</td>
<td>12</td>
<td>2</td>
<td>83</td>
<td>12</td>
<td>1</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>

TABLE 2:7

Syntactic Distribution of Initial-Mention Referents in Indeterminate Encodings

For examples, see S2.32.
Now, as is clear from what has been discussed so far, for a singular initial-mention referent in discourse, the speaker obviously has three options of morphological encoding, definite, indefinite, and indeterminate. What I am interested in now is the answer to the question what are the factors underlying the choices on the part of the speakers between the explicit, both definite and indefinite, encodings on the one hand, and the unmarked encodings on the other hand, for the referents.

Again, the investigation shows that the semantic properties of the referents play a major role here in the choice on the part of the speakers for one or the other major category of encoding. I have examined all the tokens of the bare NPs in the data. Table 2:8 compares how the singular referents in each semantic category, which are not accompanied by numerals, are distributed between the determinate and indeterminate encodings:

68
Lexical Encoding

<table>
<thead>
<tr>
<th>Determine</th>
<th>71</th>
<th>12</th>
<th>33</th>
<th>27</th>
<th>2</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(98%)</td>
<td>(86%)</td>
<td>(51%)</td>
<td>(55%)</td>
<td>(23%)</td>
<td>(96%)</td>
</tr>
<tr>
<td>(definite)</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>(indefinite)</td>
<td>71</td>
<td>12</td>
<td>24</td>
<td>23</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

| Indeterminate | 1 | 2 | 31 | 22 | 7 | 1 |
|               | (2%) | (14%) | (49%) | (45%) | (78%) | (4%) |

Total: 72 14 63 49 9 23

| Story Character | Non-human Animate | Movable Vegetation | Inanimate Object | Body Part | Clothes |

TABLE 2:8

Distribution of Initial Mention Unquantified Singular NPs in Each Semantic Category Between Determine and Indeterminate Encodings

As mentioned above, Du Bois (1980) distinguishes the intrinsic saliency and plot saliency of referents in the discourse. He claims if an object has high intrinsic saliency, it is likely to be mentioned when it first appears during its critical introduction period. If it has high plot saliency but low intrinsic saliency, it is likely to be first mentioned wherever it fits into the plot. If this happens to come early, during its critical introduction period, it will be indefinite; if not, it will usually be definite (cf. Du Bois 1980:249). The data as presented in Table 2:8 suggests that the choice between explicit and indeterminate encodings is more closely related to the saliency of the referents: Referents of either high intrinsic saliency or high plot saliency are more likely to receive determinate lexical encoding upon initial
mention. The referents with high intrinsic saliency tend to receive determinate encodings while those which are both low in intrinsic saliency and low in plot saliency. The human beings and non-human animals are higher in intrinsic saliency than inanimates. As expected, the great majority of them bear explicit encodings in the data. The referents with low intrinsic saliency are also more likely to receive determinate encodings if they stand high in plot saliency. The fact is most clearly brought out through a comparison of the NPs of body parts and those of clothes. While both categories are as a rule low in inherent saliency, in the pear stories the NPs of body parts are lower in plot saliency than the clothes terms such as the hat, apron, and so on. The contrast between the two semantic categories is obvious with the regard to the lexical encoding types upon initial mention. Out of nine body part names, only two receive determinate encodings while twenty-four out of twenty-five clothes terms are put in determinate terms, definite or indefinite.

In this section, I have examined the characteristics of the initial-mention referents that receive indeterminate encoding in the pear stories. I claim that indeterminate encoding in Chinese tends to be chosen for the initial-mention referents that are low in saliency, and determinate encoding, both definite and indefinite, for those high in plot saliency, or in intrinsic saliency. Later in Chapter 4, I will demonstrate that the distinction between high vs. low saliency not only plays a role in the encoding of the referent on its initial mention, but also bears upon different anaphoric choices later in the discourse.
2.5 **Indefinite Encodings for Non-initial Mention Referents**

2.5.1 **Preliminary**

Given the assumption that speakers have reasons to assume identifiability for the referents that have already been introduced into the discourse, these referents should be traced later through the discourse in the form of definite references. In the data, however, I find that this is not always the case. Instances are found in which non-initial mentions of referents are presented as nonidentifiable in spite of their earlier introductions. Upon examining all these instances, I conclude that they can be classified into two groups according to the factors that are identified as responsible for the phenomenon. The first group is subsumed under the category of remedial introduction, and the second is attributed to the switch of perspectives.

2.5.2 **Remedial Reintroduction**

Generally speaking, two reasons can be identified for the speakers to go through the introduction for the second time. The first is related to the false assumption held by the speaker of the identifiability of the referent. As discussed in §2.42, it happens sometimes that the initial mention referent is inappropriately encoded in definite terms when the speaker has no reason to assume the identifiability of the
referent. In most such cases, the speakers soon realize the blunder, and reintroduce the referent, this time as nonidentifiable. Consider the following example:

(1) a. Na ta jiu ba maozi na qu huan ta......
then he then BA cap take go return he

b. Dangzhong you yi ge nanhaizi meanwhile exist one CL boyhuilai de shihou, return NOM time

c. juishi na maozi gei ta na ge nanhaizi i.e. take cap give he that CL boy
huilai de shihou,
return NOM time

d. Q jiu you na le ta de san ge bala. then again take PFV he NOM three CL pear
Furthermore, a related factor is that, after speakers have introduced the new referent into the discourse as nonidentifiable, they may feel unsatisfied with the wording of the initial introduction, or feel that the presentation is not impressive enough to attract the attention of the hearer. As a remedial means, they reintroduce the referent as if it had not occurred before. In cases like this, speakers often formulate the introduction in a way that is slightly different from the first time, giving more appropriate, or more elaborated, identification for the referent. (2) and (3) below are two examples illustrating the situation:

72
(2) a. You **yī** ge **ren**..., **yī** ge **muyangren**, have one CL person one CL shepherd

b. Ø qian zhe **yī** tou yang. lead DUR one CL goat

a. "There was a person, a shepherd,

b. (who) was leading a goat."

(S-11)

(3) a. Limian you **yī** ge **guoshu** de **ren**, inside have one CL fruit:tree NOM person

b. you **ge** **ren**. have CL person

a."There was a fruit tree person inside,

b. there was a person."

(S-8)

While **muyangren** 'shepherd' is an elaboration for **yī** ge **ren** 'a person',
**ge** **ren** 'a person' is the form to replace the infelicitous **yī** ge **guoshu**
**de** **ren** 'a fruit tree person'.

The second is due to the fact that the speaker has introduced the
referent into the discourse ahead of its appropriate turn in the story.
After the introduction, the speaker returns to the natural line of the
narration, thus ignoring the newly-introduced referent. A case similar
to this is one in which, although the referent is introduced at its
turn, the narrator sidetracked to something else that was not directly
relevant to the referent. The consequence for these deviations is that
there is a distance between the initial introduction and the appropriate
place at which to add further information to the referent. In such
cases, speakers often go through the introduction once again to
facilitate the referent precessing on the part of the hearers. (4) is an
illustrative example:
(3) a. You yi ge ren uh, have one CL person uh

b. zhe jiu shi gangcai this just be just:now

women tingdao ne ge yang jiao de shengyin. we hear that CL goat bleat NOM sound

c. You yi ge ren, uh have one CL person uh

d. qian zhe yi tou yang zou guoqu. lead DUR one CL goat walk past

a. "There was person, uh,
b. this was the bleat we heard just now.
c. There was a person, uh,
b. (who) was leading a goat and walked past."

(5-4)

2.53 Switch of Perspectives

The second factor contributing indefinite noninitial mentions involves the perspective from which the speaker narrates the story. As proposed in Chafe (1979), speakers have the choice between two perspectives, or two worlds, of experience. One is the "inside" perspective or the "film" world, which describes the events as if the speakers were in the film. The other is the "outside" perspective or the "real" world, which describes the events as an outside observer. The conditions for assumed identifiability often fail to work as one switches from one world to the other. Consider the following sentences:

74
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Nevertheless, the non-initial mention referents bearing indefinite encodings as discussed above constitute only a very small percentage of all the referents that have already been introduced into the discourse (six tokens in the first group, and two in the second group). The great majority of the non-initial mention referents are encoded in terms of one anaphoric type or another.

2.6 Summary

In this chapter, I have analyzed the data to reveal how the referents are introduced into discourse, and how the three major groups of formal lexical encoding for the initial-mention referents, namely, definite, indefinite, and indeterminate, are correlated with various syntactic, semantic, and discourse-pragmatic features. In the next chapter, I will offer a brief discussion of how narrative discourse is organized in Chinese. The discussion will provide a framework in which to explicate how various anaphoric devices are chosen to track these particular referents throughout the discourse after their introduction.
Notes to Chapter 2

1 The code here refers to the sequence of the pear story in the corpus of twenty stories. (S-10) means the tenth story.

2 For any NP with only a classifier, yi 'one' can always be inserted between the noun and the classifier.

3 The reason for the neutralization can be attributed to what Givon (1984:413) has recognized as the peculiar intermediate position that it occupies on the continuum of identifiability and referentiality: "On the one hand, [generic nominals] clearly do not refer to specific individuals within the universe of tokens, and thus share some of the properties of non-referentials. On the other hand, they may be particularly when used as subjects — highly topical and continuous, and thus likely to have been entered into the active discourse file. And in this sense they share many properties of definites [identifiables]."

4 The pivotal position refers to the interverbal slot V V as exemplified by the following sentence:
(1). Ta kandao ne san ge xiaohai zhengzai zou guolai.  
He see that three CL kid DUR walk come  
"He saw the three kids coming."  (S-7)

5 Notice that which of the two handles of the bike is referred to is not important here, as explained by the Curiosity Principle expounded in Du Bois (1980).

6 The other group consists of objects that have been deactivated from an earlier active state (cf. Chafe 1984:16).

7 Two major discourse modes, narrative and descriptive, are generally recognized in the linguistic literature for clauses according to their contribution towards the advance of the discourse. The clauses in narrative mode advance the story line through presenting the major events of the discourse typically in a temporal sequence, while those in descriptive mode mainly provide background information for the major events and the participants involved. Cf. Du Bois 1980, Hopper & Thompson 1980, etc.
Chapter 3

Structure of Narrative Discourse

3.1 Preliminary

After the referents have been introduced into the narrative discourse, they become an organic part of it, and are involved in the deployment of the whole narration. As a rule, they are encoded in terms of various anaphoric devices when later mentions are made of them throughout the discourse. As referent-introduction and referent-tracking are processes that are by their nature conducted not within isolated sentences but in coherent discourse, the principles that underlie the mechanisms of the processes are to be understood only in connection with the organization of the discourse as a whole. As a result, before we probe into the issue how various anaphoric devices are used in Chinese narrative discourse, it is necessary to enquire how narrative discourse is organized so that we can establish a framework within which to investigate the use of anaphora.
I will start with the features of the narrative as seen in the taxonomy of various types of discourse. After that I will discuss the structure of narrative discourse in detail.

3.2 Taxonomy of Discourse

In accordance with Longacre’s pioneering studies on discourse taxonomies (Longacre 1976, 1983), all possible discourses can be broadly grouped into four major categories along two parameters: contingent succession and agent orientation. Contingent succession refers to a framework of temporal succession in which some (often most) of the events or doings are contingent on previous events or doings, while agent orientation refers to orientation towards agents with at least a partial identity of agent reference running through the discourse. The two parameters intersect so as to present a four-way classification of discourse types as represented in the following diagram:
Discourse Type

<table>
<thead>
<tr>
<th></th>
<th>Narrative</th>
<th>Procedural</th>
<th>Behavioral</th>
<th>Expository</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contingent</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Succession</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agent</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Orientation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DIAGRAM 3:1
Discourse Taxonomy

Of course, finer distinctions can be made if more parameters are added such as projection, tense, etc. For our purpose in the paper, however, the four-way classification is sufficient as it is.

The narrative type of discourse is chosen as the discourse within which to investigate how referents are introduced and tracked later on. In my data corpus, some narrative discourses are found to be embedded in a given conversation. To the extent that they are narratives, what is proposed below with regard to the organization of narrative discourse applies to them as well.
3.3 Structure of Narrative Discourse

3.3.1 Preliminary

It is commonly agreed that discourse is structured. Yet how discourse is organized remains a fascinating topic challenging all those who are attracted to the issue from a great variety of disciplines, including linguistics, cognitive psychology, sociology, and computer science. Quite a few theories of discourse structure have been proposed in the literature (Grimes 1975, Van Dijk 1972, 1977, Longacre 1977, 1979, 1983, Chafe 1979, 1980, Mann and Thompson 1983, 1984, 1985, Fox 1984, inter alia), which have offered invaluable stimulus to the present research. For my purpose, among all the proposals in the field, I am particularly interested in the Grammar of Discourse as framed in Grimes (1975) and Longacre (1979, 1983), the Flow Model of Discourse as proposed in Chafe (1979), and the Rhetorical Structure Theory developed by Man and Thompson (1983, 1984, 1985) and Fox (1984). Since it is not my purpose in this paper to initiate an alternative to the current theories of discourse, I will borrow freely from what has already been offered to construct a framework within to explore the use of anaphoric devices in Chinese.

Narrative discourse, like other types of discourse, is organized along two parameters, linear and hierarchical. Previous investigations (Hinds 1978, Clancy 1980, Tomlin 1981, 1984, Givón 1983, Fox 1984, inter alia) have indicated that anaphoric choice is closely related to
the position of referents in both types of organization, although more recent studies such as Fox (1984) and Tomlin (1984) argue that the hierarchical organization may play a more fundamental role in the whole process. I will return to this point in the next chapter.

As linguistic verbalization of cognitive contents, all narrative discourses can be analyzed along two intimately related levels, the cognitive and the linguistic level. Most of the studies in discourse grammar center around the identification of component units at each level of representation, the characterization of their organization, and the specification of the intertwining relations between the two representations. In accordance with Chafe (1979), the two levels of narrative discourse display the following units:

**Cognitive Level:** Focus —> Thought —> Episode

**Linguistic Level:** Phrase —> Sentence —> Paragraph

**DIAGRAM 3:2**

Units in Two Levels of Narrative Discourse

What I am going to do next is to explicate the properties of the units at each level and their interrelationships within the framework formulated by Chafe (1977, 1979), Longacre (1977, 1983), Mann and Thompson (1983, 1984, 1985), and Fox (1984). First I am going to discuss the cognitive level. Following will be a discussion of the
linguistic level, in which Chinese-specific questions will also be addressed.

3.32 Cognitive Level of Narrative Discourse

Among the four units at the cognitive level proposed by Chafe (1979), foci are the basic units of information in storage in mind in the sense that they represent the amount of information to which a person can devote his central attention at any one time. Linguistically, they are expressed in phrases, which are usually, though not always, clauses of single case frames.

Foci are clustered into thoughts in accordance with some principles of coherence. The clustering is partly a state of how foci are related to each other when in storage in mind and partly an on-line process that develops itself when the foci are verbalized. As a result, the number of foci that are manifested in a thought depends to some extent upon the speaker and cannot be predetermined. Linguistically, thought usually assumes the form of sentences, which are syntactically closed, with a characteristic sentence-final intonation contour.

The next unit higher in the hierarchy is the episode, which, as Chafe (1979) argues, refers to chunks of foci or thoughts that are marked by more or less coherent spacial orientation, coherent temporal
continuities, coherent configurations of characters, coherent event sequences, and coherent worlds. Episode boundaries are those points where there are radical changes in spatial, temporal continuity, configuration of characters, event structures, or worlds. As these features don’t always co-vary in discourse, it is impossible to divide a specific narrative unambiguously into a fixed number of episodes. The presence vs. absence of episode boundary constitutes a continuum, along which different scales can be recorded according to how many changing features are present and how radically the changes are. An episode boundary is strongly present only where all the features change in a maximal way. I will return to this point in Chapter 4 in connection with the discussion of the use of NA in discourse.
3.33 Linguistic Level of Narrative Discourse

3.331 Preliminary

At the linguistic level, the verbalization of narrative discourse is realized in terms of three linguistic units, clause, sentence, and paragraph, organized into a hierarchy. In this subsection, I will examine the syntactic aspect of the three constituent units in Chinese. After that, I will discuss the rhetorical organization of these linguistic units in discourse.

3.332 Syntactic Organization

I start with the assumption that, as in English, focus defined by Chafe (1979) as a cognitive unit usually appears as a clause in Chinese, which is defined as a single case frame, and characterized by a phrase intonation contour. In the same vein, a thought appears as a syntactically closed sentence, which is characterized by a sentence-final intonation contour in spoken Chinese, and by a sentence-final punctuation mark in written Chinese, such as a period, question or exclamation mark. Paragraph boundaries are characterized by major hesitations in spoken register and by indentation in writing. Chinese and English have no paragraph indicator as found in languages like Huichol, Shipibo etc. reported in Longacre (1979). Moreover, there is a lack of covariances of the parameters that determine the
segmentation of episodes as discussed earlier. As a result, it is impossible to impose on the speech flow a cutting point for paragraph boundaries. I will return to this point in Chapter 4 in connection with the discussion of NA in the pear stories.

After a brief account of the manifestations of clausehood, sentencehood, and paragraphhood in Chinese, I proceed to examine the constitution of the single-case-frame clause and sentence in Chinese.

The clause under discussion here is composed of a predicate and a number of entities that bear a variety of roles to the predicate which are generally termed as "cases" in the literature (cf. Chafe 1970, Fillmore 1968, 1977, Anderson 1971, Dik 1980, Longacre 1983, Foley and Van Valin 1984, etc). A distinction is usually made between participant or central case roles, such as agent, patient, recipient, instrument, and circumstantial case roles, such as location, time, extent, manner. Although all linguistic entities that assume these case roles in the clauses may receive anaphoric encodings, only those of the participant roles will be our main concern in later discussion.

In a prototypical clause, the verb predicate and their arguments assume the following format:
One-argument Verb:   NP  V

Two-argument Verb: NP  V  NP

Three-argument Verb: NP  V  NP1  NP2

DIAGRAM 3:3

It has been one of the most controversial issues in Chinese linguistics how to assign syntactic names such as subject, object, to the NPs in the various slots as shown above. Lu (1955) gives a detailed and penetrating discussion of the difficulty involved in the analysis of this grammatical problem. Take "subject" for instance. As has been stated in Keenan (1976), it tends to be a grammaticalized entity which incorporates factors of nominative case marking, semantic agentive status, and discourse topicality. These factors, however, don't always converge. In case of conflicting parameters, the question arises of how the subject is to be identified. In English, it is comparatively simple, though not without problems. Chinese, on the other hand, does not have a "subject" category which controls concord or other grammatical processes in the same way as in English, as is discussed in Li and Thompson (1976). For the present, we have no answers to the question whether there may be evidence other than what is available for languages like English supporting the existence of subject as a grammatical entity in Chinese which has not been found yet, or whether the term represents nothing more than an unwarranted
concept for this language. I decide to follow the practice adopted in Li and Thompson (1981) when mentions are made to these arguments in a Chinese clause. The preverbal NP will be called subject, postverbal NP of the two-argument verb, object, and the postverbal NP1 and NP2, indirect object and direct object respectively. As in Li and Thompson (1981), I use these terms just for easy reference to the constituents that occupy these syntactic slots in the prototypical Chinese clauses, without committing myself to a precise definition of these terms for Chinese linguistics.

Clauses concatenate into sentences. From the syntactic perspective, clauses are considered to be subordinate when they are embedded in another clause as the subject, object, complement, or attitudinal. In the hierarchy of the constituent units, these embedded subordinate clauses are not on the same level as the main clause. I will return to this point later in Chapter 4.

After the above discussion of the syntactical aspect of the three discourse units in Chinese, we move to the rhetorical organization of these linguistic units.

3.3.3 Rhetorical Organization

The term "rhetorical organization" is used to specify the cognitive relationship between the linguistic units like clauses, sentences, as discussed above. The major part of the following

89

Here, we maintain that the coherence among the linguistic units is achieved through the existence of various relational predicates, which, put in simple terms, specify the rhetorical relations among the units that comprise a larger unit. As discussed in Grimes (1975) and Longacre (1979, 1983), relational predicates are posited as higher-order predicates in discourse that relate to their constituent arguments in much the same way as verbal predicates do to their arguments. These higher-order predicates take propositions as their basic arguments, which usually assume the syntactic form of clause. In addition to single clauses, the relational predicates can also take as constituent arguments those complexes of clauses which are themselves organized in the form of a relational predicate, resulting in a discourse characterized by a hierarchical organization of the units of various size involved.

Before we enumerate the relational predicates posited below, let’s state briefly the following three major distinct, yet mutually connected, axes, along which we recognize and group these relational predicates:

- Formal parameter
- Structural parameter
- Semantic parameter
By "formal parameter" I am referring to whether there are formal lexical devices to indicate the intended relations or not; we distinguish between explicit and implicit relational predicates. For example, in the following sentences, (1) is an instantiation of an explicit relational predicate, while (1)', an implicit one, although the constituent clauses bear the same semantic relation in the two cases:

(1) a. Yinwei tian tai re, because sky too hot
   b. ta jintian meiyou chuqu. he today not go:out

   a. "Because it was too hot,
   b. he didn't go out today."

(1)' a. Tian tai re, sky too hot
   b. ta jintian meiyou chuqu. he today not go:out

   a. "It was too hot,
   b. he didn't go out today."

Usually the relational predicates impose a rather rigid order among the constituent parts if the relational content is not conveyed in terms of explicit lexical items. This restriction relaxes gradually when the constituents are themselves composed of relational predicates and found at the higher levels of the hierarchical structure of the discourse. We will return to this point later.
The "structural parameter" has to do with the structural configurations that characterize the higher-order predicate and its arguments; these relational predicates can be seen to fall into two major groups: conjoining and adjoining. In a conjoining relational predicate, the constituents are equally related, while the constituents of an adjoining predicate are unequal: one of them is the nucleus of the predicate, to which the other parts are subordinate as adjuncts. We borrow from Mann (1982), and Fox (1984) the following diagram to illustrate the structural configurations discussed here:
By "semantic parameter", I am referring to the conceptual content that characterizes the relations among the constituents of the predicates, after which the various relational predicates are named. For instance, we identify a **Condition Predicate** in the following combination of two predications:

(2) a. Yaoshi bu xiyu, 
    if not rain 
    b. wo mingtian jiu dongsheng. 
    I tomorrow then leave 

a. "If it doesn't rain, 
b. I will leave tomorrow."

\[ \text{Condition} \quad \text{Predicate} \]

\[ a \quad b \]

where (2)a. serves as the condition for the realization of (2)b. Later, we are going to present a list of major relational predicates that have been established in this way.

Disagreements among linguists with regard to the definition and recognition of the rhetorical structures in discourse have led to proposals of different numbers and descriptions of relational predicates; a detailed review of them is out of place here. What we
will enumerate below is only a list of major relational predicates that are generally recognized in the literature.

**List of Relational Predicates**

Here we present a brief account of the relational predicates that we have recognized in the rhetorical organization of linguistic units in discourse. The first four are conjoining predicates, while the rest are adjoining ones.

1. Joint Predicate

This predicate takes two or more conjoined propositions which can be taken as members in a set that are equally joined to each other. As compared with other high-order predicates, the joint predicate maintains a rather loose relationship among the constituents. The order among the members can usually be reversed without affecting the meaning of the text. Typically, the constituents are in the same mode, such as descriptive, narrative, and so on. It is one of the most basic organizing unit in the rhetorical structure of discourse and occurs at various levels of the hierarchy. Following is an example illustrating the predicate:
(3) a. You lu se,  
    have green color  

b. haoxiang you shan a, shu,  
    seem have hill PAL tree  

c. ranhou hai you huang se de daolu.  
    then still have yellow color NOM road  

a. "There was foliage,  
b. and it seemed there were hills, and trees,  
c. and then there was a yellow road."  
(S-6)

Joint

a b c

2. Contrast Predicate

The constituents subsumed in this predicate basically share the same relationship as those in Joint predicate. The only difference is the presence of contrast or antithesis among the constituents here. Following is an example illustrating the predicate:
(4) a. Ta bu shi ne ge zai neli tou chi ji ge,
   he not be that at there steal eat some CL

   b. ta shi ba ne ge yi lou a,... quanbu dai zou.
   he be BA that CL one CL PAR all take away

   a. "He didn't merely steal some to eat there,
   b. instead, he wanted to take away the whole basket."
   (5-11)

Contrast

a

b
3. Alternation Predicate

This predicate contains two or more conjoined constituents from which either one may be chosen. Following is an example illustrating the predicate:

(5) a. Yaomo wo qu,  
either I go

b. yaomo ni qu.  
or you go

a. "Either I go,
b. or you go."

4. Succession Predicate

The constituents in this predicate are bound to each other through a chronological or logical sequence among themselves. They usually cannot shift their order without changing the meaning. Following is an example illustrating the predicate:
(6) a. Ta ting xiala,
    he stop down

b. G ba yi lou shuiguo tongtong ban dao che shang,
    BA one CL fruit all move to bike up

c. ranhou G zou le.
    then go CRS

a. "He stopped,
b. moved a whole basketful of fruit upon the bike,
c. and left."

(S-20)

5. Background Predicate

The predicate is composed of two or more constituents. One of
them is the nucleus, to which the others are adjoined, providing the
local, temporal, or event background. Following is an illustrative
example:
(7) a. Ta yi zhan qilai de shihou,
   he once stand up   NOM time
   
   b. 'kandao san ge nanhaizi zai ta qiannian a.
       see three CL boy   in he front   PAR

   a. "When he stood up,
   b. (he) saw three boys in front of him."
   (S-16)

6. Elaboration Predicate

The relational predicate contains one nucleus and one optional
adjunct. Between the nucleus and the adjunct stands a variety of
conceptual relations, all of which are subsumed under the elaboration
upon a proposition or complex of propositions. As summarized in Mann
(1982), the adjuncts elaborate on the nucleus by offering various
kinds of details, including:

    set: member
    abstraction: instance
    whole: part
    process: step
    object: attribute

The elaboration can go both ways with the above-mentioned
relationships. The adjunct may add more specific information to a
generic statement, or add a more generic information to a specific statement, as exemplified by the following instances:

(8) a. He was executed yesterday.

   b. He was shot by the firing squad.

   a. set:   b. member
     execute     shoot

   Elaboration

   a   b

(9) a. They dug up Assyrians.

   b. They did some excavation.

   a. instance:    b. abstraction
     dug up Assyrians   do some excavation

   Elaboration

   a   b

7. Reason Predicate

The nucleus of this relational predicate has an adjunct that indicates the
reason for it. The term "reason" is used here in a rather
comprehensive way, including what are often identified as cause,
evidence, explanation, motivation, etc., elsewhere in the literature.

Following is an illustrative example:

(10) a. Yinwei ta tui ye die shang le,  
    because he leg also fall wounded CRS

    b. ta meiyou qi shangqu.  
    he not ride up

a. "Because he got wounded in the leg in the fall,
b. he didn’t get on the bike."  
   (5-7)

8. Concession Predicate

The adjunct in the predicate represents a potential refutation to
the nucleus. The addressee doesn’t challenge the validity of the
adjunct proposition, but asserts the co-occurrence of the nuclear
proposition and the adjunct proposition that is contrary to
expectation. Following is an illustrating example:
9. Condition Predicate

The adjunct here states the condition that must hold for the nuclear proposition to follow. Following is an illustrating example:

(12) a. Ruguo ne ge shi ge daren dehua,
    if that CL be CL adult if

    b. wo hui bu yuanliang ta.
    I will not forgive he

    a. "If he were an adult,
    b. I couldn’t forgive him."
    (S-17)
For each of the relational predicates listed above, there are a large number of lexical expressions in Chinese that are used to indicate explicitly the relationships among the constituents. For a list of these connectives and adverbials, see Lü (1981), and Wu (1982).

In many cases, however, the specific conceptual relations characteristic of these relational predicates are understood through inference on the basis of the linguistic and extra-linguistic knowledge. Among the linguistic devices, the relative order of the propositions involved offers crucial information as to the interpropositional relations. Thus it may be vital to the success of communication that, for some of these relational predicates, their clauses should have a fixed order in the absence of explicit lexical items specifying their relationship. For instance, in the absence of explicit connective expressions, the nucleus-preceding-adjunct order in the Elaboration Predicate, and adjunct-preceding-nucleus order in the Reason, Concession, and Condition Predicates are crucial for a clear indication of nucleus vs. adjunct status of the propositions subsumed within the relational predicate, although further information in the discourse is necessary to determine precisely what kind of
specific inter-clausal relation is intended in a particular case. On the other hand, if there are such connectives present, the above-mentioned orders are as rigid as otherwise.

3.4 Summary

In this chapter, we have discussed briefly the structure of narrative discourse in Chinese in terms of its two levels of representation, the constituent units at each level, and the interrelationships among the units. This offers a framework with which we can investigate the use of various anaphoric devices that track referents throughout discourse.
CHAPTER 4

Referent Tracking in Chinese Narrative Discourse

4.1 Preliminary

Within the framework of the grammar of discourse as proposed in Chapter 3, we are going to examine in detail the use of various anaphoric devices which are chosen to track the various types of referents after they have been introduced into discourse. I will begin the investigation with a brief discussion of the three major anaphoric types in Chinese. Following will be a general picture of the distribution of the three anaphoric devices from discourse-pragmatic, syntactic, and semantic perspectives. After that, we will develop an in-depth investigation of the use of each of the three major anaphoric types that will account for both their general distribution and their specific usage.
4.2 Anaphoric Types in Chinese

The anaphoric devices in Chinese broadly fall into three formal categories: nominal, pronominal, and zero anaphora. Let's take each of them in turn.

4.21 Nominal Anaphora (NA)

Typically, a nominal anaphor stands in a coreferential relationship with its antecedent. Halliday and Hasan (1976:314) maintain that a more precise characterization for the relation between the anaphora and the referent involved would be co-interpretation, which means that the interpretation of the former depends in some way on that of the latter, although they may not necessarily have the same referent. From this perspective, coreference is only one particular form that co-interpretation may take. Consider the following sentences:

(1) a. — Arthur's very proud of his chihuahuas.
    b. -- I don't like them.

(1) b. can mean "I don't like Arthur's chihuahuas" or "I don't like chihuahuas". Them in the first interpretation is coreferential with Arthur's chihuahuas, but not in the second interpretation, in which case it is to be explained as being coreferential with the noun head

106
chihuahuas taken on its own. In our data, we are only concerned with the anaphors that show identity of reference to their antecedents. Semantically, the noun phrases used in this way are of identifiable reference by definition. Formally, they assume various lexical encodings, including bare nouns, possessed NPs, and NPs preceded by various determiners, including demonstratives, numerals, and so on. The most commonly used determiners to make anaphoric reference are demonstratives like zhe and ne and their plural forms zhexie and nexie. (When followed by numerals, however, the plural forms are replaced by the singular forms.) Consider the following examples:

(2) a. Ta jiu gei na ge huan ta maozi de xiaohaizi he the give that CL return he cap NOM kid
    san ge bala.
    three CL pear

b. Ranhou, ne ge xiaohaizi jiu ba san ge bala then that CL kid then BA three CL pear
    fen gei ta.
    divide give he

a. "Then he gave three pears to the kid who returned the cap to him,
b. Then, the kid gave him three pears."

(S-10)
(3) a. Nanhaizi ...... na le san ge shuigu,  
    boy take PFV three CL fruit  

b.  tgg jiaogei tgg le zhe san ge haizi.  
    give PFV this three CL kid  

c.  Zhe san ge haizi hen gaoxin.  
    this three CL kid very pleased  

a. "The boy ... took three pieces of fruit,  
    b. (he) gave (them) to the three kids.  
    c. The three kids were very pleased."  

(S-14)

Following is an example illustrating other types of NA, a noun preceded by a numeral in (4)a., and a bare noun in (4)b.:

(4) a. San ge ren wang na ge nongren de na ge  
    three CL person to that CL farmer NOM that CL  
    fangxiang zou.  
    direction walk  

b. Neme, nongren zhe ge shihou ganghao cong  
    then farmer this CL time just from  
    shu shang xialai.  
    tree up down  

a. "The threesome.....walked in the direction of  
    the farmer,  
    b. Then, the farmer just came down from the tree."  

(S-5)

4.22 Pronominal Anaphora (PA)

Chinese third-person pronouns don't differentiate sexes. Although three different characters have been introduced into the language to indicate male, female, and non-human referents in written form, in spoken Chinese it is the same ta that is equivalent to "he", "she", and "it" in English. To avoid interference from the difference between
the written and the spoken Chinese in this aspect, I have excluded from my written data those examples in which the referent of a pronominal anaphor can be differentiated from its competing candidate through the written form but not the spoken form. The plural form for pronouns is formed though adding a suffix men to ta. Consider the following examples:

(1) a. Ne ge xiaohai... qi zhe zixingche......
   that CL kid ride DUR bike

   b. Ta kandao yi ge xiaonuhai,
      he see one CL girl

   c. suoyi, ta jiu yimian zoulai.
      so she then from:opposite come

a. "The kid ... was on a bike...
  b. He saw a girl,
  c. then, she came from the opposite direction.

(S-20)

(2) a. Ganghao zhe ge shihou,
    just this CL time

   b. san ge xiaohaizi zai kan ta.
      three CL kid DUR watch he

   c. Suoyi, ramen bang ta ba bala jian qilai.
      so they help he BA pear pick up

a. "Just then,
  b. three kids were watching him,
  c. So, they helped him pick up the pears."

(S-10)
4.23 Zero Anaphora (ZA)

As explicated in Chen (1984), ZA refers to the situation in which there is a syntactic ‘hole’ in the sentence where a referent is understood but not explicitly mentioned. It is considered as one of the three referential types in that the choice between zero anaphora and the other two types is pragmatically controlled. Change in the context could bring a less attenuated referential form into the syntactic hole. Consider the following sentences:

(1) a. Duimian lai le yi ge nuhaizi,
    opposite come PFV one CL girl

    b. yu ye qi zhe zixingche.
        also ride DUR bike

    a. "Opposite came a girl,
    b. (she) was also on a bike."

(5-6)

The preverbal slot in the second clause is assumed to be an instance of ZA, which refers to ‘a girl’ in the preceding clause. The slot could be occupied by a pronominal or nominal anaphor as required by pragmatic considerations without any changes to the propositional content of the sentence, as shown in (2) and (3):

(2) a. Duimian lai le yi ge nuhaizi,
    opposite come PFV one CL girl

    b. ta ye qi zhe zixingche.
        she also ride DUR bike

    a. "Opposite came a girl,

    110
b. she was also on a bike."

(3) a. Du miaan lai le yi ge nuhaizi, opposite come PVF one CL girl

b. zhe ge nuhaizi ye qi zhe zixingche. this CL girl also ride DUR bike

a. "Opposite came a girl.

b. the girl was also on a bike."

In the Chomskyan Government and Binding paradigm, a PRO with the feature matrix [+referential, -overt] is recognized in the following sentence (cf. Chomsky 1981, Aoun 1985):

(4) Lao Zhang xiang [\( \texttt{\_g} \) PRO qu].

Lao Zhang want \( \texttt{\_g} \) go

"Lao Zhang wants \( \texttt{\_g} \) to go."

The PRO is taken to be the phonetically unrealized counterpart of phonetically realized lexical name or pronoun. Our concern in the present study, however, is the pragmatically controlled ZA, as exemplified in (1)b. It can be differentiated from instances of syntactic null elements by a simple test (cf. Bosch 1983): A zero anaphor may be replaced by a nominal or pronominal expression that refers to the same referent in the context at hand but not necessarily of the same form as the antecedent. For instance, consider the following sentence:

111
(5) a. Lao Zhang hen yuanyi qu,
Lao Zhang very wish go

b. danshi  a you bu ken shuo chulai.
but yet not willing speak out

"Lao Zhang wants to go very much,
but (he) won’t speak out."

As is the case with (1), the ZA in (5)b. can be replaced by expressions which are different from the antecedent in form but not in reference without affecting the acceptability of the sentence. Examples are (6) and (7), where ta ‘he’ and zhe wei laoxiong ‘this old chap’ respectively take the place originally occupied by zero anaphora:

(6) a. Lao Zhang hen yuanyi qu,
Lao Zhang very wish go

b. danshi ta you bu ken shuo chulai.
but he yet not willing speak out

a. "Lao Zhang wants to go very much,
b. but he won’t speak out."

(7) a. Lao Zhang hen yuanyi qu,
Lao Zhang very wish go

b. danshi  zhe wei laoxiong you bu ken shuo chulai.
but this CL chap yet not willing speak out

a. "Lao Zhang wants to go very much,
b. but the old chap won’t speak out."

In contrast, a coreferential nominal or pronominal expression cannot take the place of the syntactically controlled null elements such as are found with an embedded verb or with the verb which comes after the first verb in a verbal serial construction. Consider the following examples:
The term "zero anaphora" only refers to those missing arguments which are pragmatically controlled as in (1) and (5), but not to the syntactically controlled null elements as in (4), (8) and (9). The rules which constrain the occurrences of syntactical null elements are beyond the scope of the present investigation.

Zero anaphora raises a problem which doesn't come up with the other two referential categories. While the semantic case roles of the missing elements can be inferred from the subcategorization rules of predicate and the concomitant arguments, it is often rather difficult
to determine exactly what syntactic role an element encoded in terms of ZA occupies in the sentence in the same way that the syntactic roles of nominal or pronominal anaphors are recognized. However, we can infer the syntactic slots that the referent in ZA occupies in a specific sentence on the basis of the formation of the prototypical Chinese sentence. As is usually accepted in functional linguistics, the main, declarative, affirmative, active sentence is taken as the prototypical sentence (cf. Givón 1979, 1984). In Chinese, the distribution of concomitant arguments of predicate verbs like song 'send', jie 'lend', and gei 'give', assume the following form in a prototypical sentence in the same order as in English:

Preverbal NP V Postverbal NP1 Postverbal NP2

Prototypically, Preverbal is the slot for topic/subject, and Postverbal NP1 is the slot for direct object. On the other hand, based on cross-linguistic evidence, Givón (1984) arranges the major semantic case roles into the following hierarchy of access to topic/subject:

AGT > DAT/BEN > PAT > OTHERS

The higher the case role is on the hierarchy, the more likely it is to have access to the position of clausal subject/topic. As a result, the concomitant semantic case roles of AGT, DAT, and PAT assume the following order in a prototypical Chinese clause:

114
(10) Ne ge xaiohai ge le tamen san ge bala.
that CL kid give PFV them three CL pear

AGT V DAT/BEN PAT

"That kid gave them three pears."

(S-7)
4.3 General Distribution of Anaphoric Types

4.31 Preliminary

It is helpful to provide a general picture that depicts how the referential forms belonging to one of the three categories are distributed in the data before I go further into details of discussion. All the tokens of the three anaphoric types in the data have been tabulated quantitatively along various parameters. The results are presented in the form of eight tables. Following is a list of them:

Table 4:1 Frequency of the three major categories of anaphors
Table 4:2 Anaphora and referential distance in terms of number of intervening clauses
Table 4:3 Anaphora and referential distance in terms of number of intervening sentences
Table 4:4 Anaphora and number of potential interfering referents
Table 4:5 Anaphora and Persistence
Table 4:6 Distribution of anaphors in various syntactic slots
Table 4:7 Distribution of anaphors in various semantic case roles
Table 4:8 Inanimate anaphors in various syntactic slots

116
The eight tables will offer an overview of how the anaphoric choices are correlated with discourse, syntactic, and semantic features in Chinese. They fall into two groups. The first five tables focus upon the discourse-pragmatic aspects of the anaphoric distribution, while the next three focus upon the syntactic and semantic aspects of the issue. The parameters adopted in Tables 4:2, 4:3, 4:4, and 4:5 are almost the same as those used in Clancy (1980), Givón (1983), Rude (1984), Chen (1986), *inter alia*. In the following discussion of each of these tables, I will compare the results obtained here with those reported in previous studies when appropriate, in the hope that they will give us an opportunity to ascertain tentatively to what extent the correlations are eligible for the status of language universals, and to what extent they are specific to the Chinese language only. The two groups of tables will be discussed in two separate sections.

4.32 Discourse-pragmatic Aspects of Anaphoric Distribution

In this section, we will concentrate on the discourse-pragmatic aspects of the anaphoric distribution as displayed in the first five tables. Next, let’s consider them by turn.

Table 4:1 presents the frequency of each of the three referential categories in the twenty pear stories.
<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>363</td>
<td>(26.5%)</td>
</tr>
<tr>
<td>Pronominal</td>
<td>490</td>
<td>(35.8%)</td>
</tr>
<tr>
<td>Zero</td>
<td>516</td>
<td>(37.7%)</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>1369</strong></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 4:1**

Frequency of the Three Major Categories of Anaphora

From Table 4:1, we can see that the three major anaphoric types in Chinese each account for roughly one-third of the total number of anaphors. Totalling 1,369 in number, 37.7 percent of them assume the form of zero anaphora. As compared with English and Japanese (Clancy 1980), this percentage is higher than that in English (20.5%), but lower than in Japanese (73.2%), which has no pronominal anaphora.

Table 4:2 and Table 4:3 show how distance to its previous mention in the discourse correlates with the referential choices.
<table>
<thead>
<tr>
<th>Number of Intervening Clause</th>
<th>0</th>
<th>1</th>
<th>2-3</th>
<th>4-5</th>
<th>6-10</th>
<th>10-20</th>
<th>20+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nominal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pronimnal</td>
<td>29(6%)</td>
<td>315(65%)</td>
<td>111(22%)</td>
<td>21(4%)</td>
<td>9(2%)</td>
<td>3(1%)</td>
<td>2</td>
</tr>
<tr>
<td>Pronimnal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total:</td>
<td>28</td>
<td>878</td>
<td>233</td>
<td>79</td>
<td>73</td>
<td>30</td>
<td>46</td>
</tr>
</tbody>
</table>

**TABLE 4:2**

Anaphora and Referential Distance in terms of Number of Intervening Clauses

119
Number of Intervening Sentences

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4-5</th>
<th>5+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>14(4%)</td>
<td>148(41%)</td>
<td>60(16%)</td>
<td>31(9%)</td>
<td>36(10%)</td>
<td>74(20%)</td>
</tr>
<tr>
<td></td>
<td>(3%)</td>
<td>(27%)</td>
<td>(69%)</td>
<td>(80%)</td>
<td>(84%)</td>
<td>(94%)</td>
</tr>
<tr>
<td>Pronominal</td>
<td>203(42%)</td>
<td>250(51%)</td>
<td>23(5%)</td>
<td>6(1%)</td>
<td>6(1%)</td>
<td>2(---)</td>
</tr>
<tr>
<td></td>
<td>(35%)</td>
<td>(45%)</td>
<td>(26%)</td>
<td>(15%)</td>
<td>(14%)</td>
<td>(3%)</td>
</tr>
<tr>
<td>Zero</td>
<td>355(69%)</td>
<td>152(29%)</td>
<td>4(1%)</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>(62%)</td>
<td>(28%)</td>
<td>(5%)</td>
<td>(5%)</td>
<td>(2%)</td>
<td>(3%)</td>
</tr>
<tr>
<td>Total:</td>
<td>572</td>
<td>550</td>
<td>87</td>
<td>39</td>
<td>43</td>
<td>78</td>
</tr>
</tbody>
</table>

TABLE 4.3

Anaphora and Referential Distance in Terms of Number of Intervening Sentences

In general, the results reported here provide supporting evidence for the claim that distance between the two mentions of a referent bears a close relationship to the choice of referential forms. Next, we’ll discuss the two tables.

The figures in Table 4:2 show that the attenuated forms (ZA or PA) tend to be used to encode referents with a short distance to their previous mention. The greater the distance, the more likely such a referent is to have more explicit form. Over half of the referents (54%) within one-clause distance to the previous mention appear as ZA. On the other hand, the percentage of nominal encodings rises steadily as the distance expands. Beyond a distance of five clauses, it is the predominant referential form found in the data. The percentage for pronominal encoding stands between the other two types along the whole
range of distance except at the both ends. It is noted that the twenty-nine pronominal anaphors which fall within the same clause as their antecedents are all in possessive form. For the time being, we may assume that the choice for pronominal anaphora here is syntactically determined and falls outside the scope of this chapter.

Table 4:3 gives the measures in terms of the sentence boundaries between two mentions of the referent. As linguistic units, clauses and sentences in Chinese and other languages such as Japanese are not always clearly distinguished from each other. In my investigation of the data, I follow an approach that was adopted for both English and Japanese in Clancy (1980), and for Chinese in Cumming (1984). Clauses are counted in terms of their case frame (where a verb, an adjective, or a noun can serve as the predicate in Chinese), while sentences are taken to be utterances characterized by specific intonational contours in speaking and by a period in writing.

The general picture of the distribution of the referential forms in terms of sentential distance between two mentions of the referent is very similar to that given in Table 4:2 in that the more attenuated forms occur within shorter distance to the previous mention. When mention was made of a referent within the same sentence, over half (69%) of the anaphors assume zero encoding and 35% of them pronominal encoding. Within one-sentence distance to the previous mention, about half (45%) of the anaphors appear as PA. Beyond that distance, NA is the overwhelming encoding type for the anaphors. Furthermore, Table 4:3
demonstrates that over half (70%) of the zero anaphors occur within the same sentence as their co-referents, and about half (51%) of the PA are separated from their co-referents by one-sentence boundary. Also, while the NA is more evenly distributed along the scale of distance, 41% of the tokens are only one sentence away from their antecedents. The statistics here demonstrate that PA and NA are correlated with the sentence boundary to a certain extent. The fact that this correlation is not very strong suggests that there are other factors yet to be determined that undermine the correlation.

Next, let's consider the relation between anaphoric choices and the number of potential interfering referents. In assessing this parameter, a referent is considered to be potentially interfering to the anaphoric reference if it meets the following three conditions:

1. It occurs between the anaphora and its antecedent in the preceding discourse.

2. It is not ruled out as a candidate for the reference of the anaphora by the selectional features of the predicate and the arguments involved.

3. It assumes a non-zero form. A referent in ZA is not considered to be interfering even if it fulfills the first two conditions.
Table 4:4 shows how the number of potentially interfering referents correlates with referential choices.

<table>
<thead>
<tr>
<th>Number of Potential Interfering Referents</th>
<th>0</th>
<th>1</th>
<th>2-4</th>
<th>5+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>42</td>
<td>166</td>
<td>132</td>
<td>23</td>
</tr>
<tr>
<td>(5%)</td>
<td>(56%)</td>
<td>(92%)</td>
<td>(100%)</td>
<td></td>
</tr>
<tr>
<td>Pronominal</td>
<td>379</td>
<td>102</td>
<td>9</td>
<td>--</td>
</tr>
<tr>
<td>(42%)</td>
<td>(34%)</td>
<td>(6%)</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>Zero</td>
<td>473</td>
<td>41</td>
<td>2</td>
<td>--</td>
</tr>
<tr>
<td>(53%)</td>
<td>(10%)</td>
<td>(2%)</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>Total:</td>
<td>894</td>
<td>299</td>
<td>143</td>
<td>23</td>
</tr>
</tbody>
</table>

TABLE 4:4
Anaphora and Number of Potential Interfering Referents

From Table 4:4 it is clear that whether there is any referent available in the context that may be a competing candidate for the intended referent of the anaphora plays an important role in determining the referential choices. Since ZA itself carries no information at all with regard to its referent, the chances are high for ambiguity if the referent is encoded in ZA in the presence of other referents which are potential candidates for the reference (except in cases to be explained later). As a result, it is easily understood that in our data the
percentage of ZA is highest when there are no interfering referents for the slot in the environment; with more competing candidates, the number of instances of ZA decreases.

NA, on the other hand, carries the most information as compared to the other two types. It is naturally the appropriate type in a situation where disambiguation is desired. As Table 4:4 shows, it registers the highest percentage in the presence of interfering referents. Beyond the number of five concomitant interfering referents, it is the sole form that is used.

It has been suggested that the parameters of distance and interference which are examined in relation to referential choices in Tables 4:2, 4:3, and 4:4 are closely related to cognitive abilities of human beings. As Clancy (1980:129) puts it:

If, for example, the human mind is capable of dealing with only a limited number of explicit and/or inexplicit referents at a time, then this limitation will surely play a part in determining the nature of the "rules" for reference in any language. So far as adult speakers successfully empathize with their addressees' needs, the trends observable in speech samples such as these narratives should reveal the range of referential forms which a listener can process. Two possible cognitive constraints on this
ability are the amount of time that has passed since the last mention of a referent and the number of other referents mentioned in that interval.

Cross-linguistic investigations have verified the correlation between these two parameters and the referential choices in discourse to a certain degree, although languages may differ with regard to the cutting-off points of encoding types or interaction of the parameters. Givón (1983) contains abundant supporting evidence from a large range of languages (including Japanese, Amharic, Ute, Biblical Hebrew, Spanish, Hausa, Chamorro, spoken and written English). The three tables discussed above demonstrate that Chinese is no exception in this respect. In the pear stories, the pear man who is introduced at the beginning of the movie appears again at the end. Between the first episode and the last, many intervening actions have taken place, and many objects introduced. When the pear man is mentioned again at the end of the story, all narrators encode him in nominal forms without exception.

While the two parameters above assess the continuity of referents to the preceding discourse, empirical investigations reported in the literature show that the persistence of the referent in the following discourse may also be an important aspect of the discourse-pragmatic information of the referent that must be taken into consideration in the study of relevant grammatical issues (cf. Givón 1983, Chen 1986). Here, the referents are also assessed for the persistence in the
following discourse in terms of the number of subsequent clauses after the anaphor in which the referent continues an uninterrupted presence as a semantic argument of the clause, an argument of whatever role and marked by whatever grammatical means. It is assumed that more important referents have a higher probability of longer persistence. Table 4:5 shows the persistence value of each of the three anaphoric types:

<table>
<thead>
<tr>
<th>Type</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>1.55</td>
</tr>
<tr>
<td>Pronominal</td>
<td>2.54</td>
</tr>
<tr>
<td>Zero</td>
<td>2.47</td>
</tr>
</tbody>
</table>

**TABLE 4:5**
Average Number of Clauses in which Anaphor Continues Uninterrupted Presence

As reported in Table 4:5, our data show that Chinese differs somewhat from Ute (Givón 1980, 1983b) or English (Brown 1983), etc. in that ZA is shorter instead of longer in persistence than PA². It agrees with the other languages in that NA is shortest in persistence. I will return to this issue later in the chapter.

Assuming for the time being that the two parameters, distance and interference, which are taken up in Tables 4:2, 4:3, and 4:4, measure the degree of continuity with the preceding discourse, or predictability, of the referents, the results demonstrate how the
functional domain of topic identification is correlated with the anaphoric choices. The relationship can be schematically presented as the following:

```
Continuous          Discontinuous
Predictable         Unpredictable
____________________
zero anaphora
____________________
pronominal anaphora
____________________
nominal anaphora
```

**DIAGRAM 4.1**

While the correlation between referential choices, on the one hand, and distance and interference, on the other hand, has been observed in Chinese as well as in other languages, recent studies suggest that distance and interference may not be the most basic factors affecting the continuity of the referents but are only derivatives of other fundamental factors which are related to the hierarchical organizations of discourse (cf. Grosz 1977, Reichman 1981, Fox 1984, Tomlin 1982, 1984). I will take up this issue later in this chapter.
4.33 **Syntactic-semantic Aspects of Anaphoric Distribution**

In the discussion of the first five tables, we have examined the discourse-pragmatic aspects of the anaphoric choice by exploring the correlation between referential distance and potential interference, on the one hand, and anaphoric distribution in the twenty pear stories, on the other hand. Below, we will examine the syntactic and semantic aspects of the anaphoric encodings. The results of our examination of the pear stories are presented in the following three tables. Now, let’s discuss them one by one.

Table 4:6 presents the percentage of each referential category in each of the six syntactic slots:
<table>
<thead>
<tr>
<th>Syntactic Slot</th>
<th>Preverbal</th>
<th>BA</th>
<th>Postverbal 1</th>
<th>Postverbal 2</th>
<th>Pivotal</th>
<th>Oblique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>191 (55%)</td>
<td>42 (12%)</td>
<td>83 (24%)</td>
<td>2 (1%)</td>
<td>10 (3%)</td>
<td>17 (5%)</td>
</tr>
<tr>
<td></td>
<td>(20%)</td>
<td>(72%)</td>
<td>(40%)</td>
<td>(100%)</td>
<td>(23%)</td>
<td>(61%)</td>
</tr>
<tr>
<td>Pronominal</td>
<td>285 (68%)</td>
<td>16 (4%)</td>
<td>73 (17%)</td>
<td>—</td>
<td>33 (8%)</td>
<td>11 (3%)</td>
</tr>
<tr>
<td></td>
<td>(30%)</td>
<td>(28%)</td>
<td>(35%)</td>
<td></td>
<td>(77%)</td>
<td>(39%)</td>
</tr>
<tr>
<td>Zero</td>
<td>462</td>
<td>—</td>
<td>53</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(50%)</td>
<td></td>
<td>(25%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total:</td>
<td>938</td>
<td>58</td>
<td>209</td>
<td>2</td>
<td>43</td>
<td>28</td>
</tr>
</tbody>
</table>

**TABLE 4:6**

Distribution of Anaphoric Forms in Various Syntactic Slots

As demonstrated by the table, the syntactic positions differ with regard to the frequencies of the referential forms that turn up. In the preverbal position, which is usually the position for topic/subject, half of the referents in the data assume zero anaphora. Next in frequency comes the PA, which registers 30% of all referential forms. Last is the NA, with only 20% of the anaphors. The situation is reversed in the postverbal 1 position, where the NA is the highest in frequency, and ZA the lowest. It must be pointed out that, as seen from the data, the postverbal 2 position is rarely used. In the twenty peer stories, only two anaphors are found in that position, both of which assume nominal encoding. In the position of the pivotal object, 77% of the referents are found in pronominal encoding, and the rest in
nominal encoding. In the position of oblique object and in the BA construction, however, we have a higher percentage of NA than PA.

What we are doing here is ascertaining the correlation between two sets of linguistic entities, anaphoric types and syntactic positions in clauses. As discussed in the preceding paragraphs, the former are assumed to be contingent upon the continuity of the referents in discourse, while the latter are syntactic facts within the domain of the clause. There is evidence, however, which indicates that the two kinds of information are intimately related, because syntactic slots arise in grammar as a response to the same sorts of communicative needs as do distinctions in coding referents. For example, Givón (1984:137-138) has suggested that the syntactic slots constitute a topic hierarchy like the following:

SUBJ > DO > OTHERS

DIAGRAM 4:2

The higher the slot is in the hierarchy, the more likely it is to encode a referent of high topicality. Below, a similar topicality hierarchy is proposed for the syntactic slots in Chinese on the basis of the evidence presented in Table 4:6:
The NP in the BA construction is grouped together with the oblique object because BA behaves like a regular preposition in many respects. Now it is easy to see the encoding range of the various syntactic slots along the dimension of topicality displayed by different anaphoric types. ZA, which is the smallest in the hierarchy of phonological size, is confined to the first two rungs of the syntactic hierarchy. Its cutting-off point lies at Postverbal 1. PA, which stands between ZA and NA in topicality, is most closely affiliated with the Postverbal 1 position, pivotal object, and oblique object slots, which all stand at the middle part of the hierarchy. The cutting-off point for pronominal anaphora is the last rung but one in the hierarchy. It is in fact "ungrammatical" for most di-transitive verbs to have a pronoun in the slot of postverbal 2. Consider the following sentences:
(1)  Wo song le ta yi ben shu.
     I give PFV he one CL book

(1)'
*Wo song le yi ben shu ta.
  I give PFV one CL book he

"I gave him a book."

Ta in (1)' must be raised in the hierarchy to the level of oblique object if it is intended that it remain in the end-focus position:

(2)  Wo song le yi ben shu gei ta.
     I give PFV one CL book give he

"I gave a book to him."

The slots that allow NA range over the whole hierarchy. The last two rungs in the hierarchy contain more than half of the referential forms encoded in NA in the data.

In summary, the more attenuated anaphoric forms tend to appear in syntactic slots higher up the hierarchy. This finding is compatible with those in §4.32, when we consider the fact that the syntactic slots are nothing but the grammaticalization of the same communicative needs that underlie the correlation between anaphoric choice and the discourse-pragmatic parameters.

Next comes the semantic aspect of the anaphoric distribution. Table 4:7 presents the correlation between the anaphoric types and various semantic case roles.
<table>
<thead>
<tr>
<th>Semantic Slot</th>
<th>AGT</th>
<th>DAT</th>
<th>PAT</th>
<th>LOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>125(36%)</td>
<td>36(11%)</td>
<td>143(41%)</td>
<td>42(12%)</td>
</tr>
<tr>
<td></td>
<td>(18%)</td>
<td>(16%)</td>
<td>(46%)</td>
<td>(70%)</td>
</tr>
<tr>
<td>Pronominal</td>
<td>191(46%)</td>
<td>114(27%)</td>
<td>95(23%)</td>
<td>18(4%)</td>
</tr>
<tr>
<td></td>
<td>(28%)</td>
<td>(505)</td>
<td>(30%)</td>
<td>(30%)</td>
</tr>
<tr>
<td>Zero</td>
<td>366(71%)</td>
<td>76(15%)</td>
<td>74(14%)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(54%)</td>
<td>(34%)</td>
<td>(24%)</td>
<td>—</td>
</tr>
<tr>
<td>TOTAL:</td>
<td>682</td>
<td>226</td>
<td>312</td>
<td>60</td>
</tr>
</tbody>
</table>

TABLE 4.7
Distribution of Anaphoric Forms in Various Semantic Slots

As Comrie (1981) and Givón (1983; 1984:139) propose, semantic case roles also form a hierarchy according to the likelihood of their coding the more continuous topic in discourse. The hierarchy is formulated as the following:

AGT > DAT > BEN > PAT > LOC

DIAGRAM 4.4

In a prototypical clause, the claim is that the higher the semantic case role is in the hierarchy, the more likely it will be the syntactic subject of the sentence.
As discussed above, the various syntactic slots correlate with anaphoric types with respect to topicality. Since the semantic case roles are ranked according to their likelihood of occupying the syntactic slot of subject, or direct object, etc., it is easily predicted that they also show a correlation with different anaphoric types along the same parameter of topicality.

The results presented in Table 4:7 verify the prediction. As shown in the table, AGT has the highest percentage (54%) of ZA, which decreases in frequency along the hierarchy until it is cut off after PAT. PA registers higher with PAT, and is rather evenly distributed with the other semantic case roles. NA has the lowest frequency among the three in the case roles of AGT and DAT, but ascends to highest in the last two case roles of the hierarchy.
Now let's discuss Table 4:8, which displays the distribution of the inanimate pronominal and nominal anaphors in various syntactic slots.

<table>
<thead>
<tr>
<th>Syntactic Slot</th>
<th>Pre-verbal</th>
<th>Post- verbal 1</th>
<th>Pivotal Object</th>
<th>Oblique Object</th>
<th>Post-verbal 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>42 (22%)</td>
<td>52 (63%)</td>
<td>3 (30%)</td>
<td>50 (91%)</td>
<td>2 (100%)</td>
</tr>
<tr>
<td>Pronominal</td>
<td>4 (1.4%)</td>
<td>--</td>
<td>2 (6%)</td>
<td>6 (37.5%)</td>
<td>--</td>
</tr>
</tbody>
</table>

**TABLE 4:8**

Inanimate Anaphors in Various Syntactic Slots

According to Comrie (1980), Givón (1984:107), animacy is one of the cluster of properties that are related to agentivity. They have proposed the following animacy hierarchy:

Human > Animate > Inanimate > Abstract

**DIAGRAM 4:5**
If this scale of animacy is also related to the likelihood of accession to topic/subject in the syntactic hierarchy, as has often been suggested, we would expect a diminishing percentage of inanimate anaphora as we ascend the syntactic hierarchy. The results in Table 4:8, however, don't provide full support for the prediction. The percentage of inanimates decreases as we go from the bottom of the hierarchy from Postverbal 2 through Pivotal Object to Subject position. All appear just as predicted except in the Postverbal 1 slot, which displays a high degree of irregularity in that the inanimate NA in this slot records a percentage that is higher than that in the next rung down the hierarchy. Moreover, no PA of inanimate reference is found in this position in the data. Although it is "grammatical" to have an inanimate pronoun in the position of Postverbal 1 as (3)b. below, sentences like this sound very artificial and stilted:

(3)  
   -a. Ni judeqi zhe kuai shitou a?  
      you can:lift this CL stone PAR
   
   -b. Wo jubaqi ta.  
      I cannot:lift it

   -a. "Can you lift the stone?"

   -b. "I cannot lift it."

It is much more natural to replace ta in (3)b. by a nominal form as in (4), or by ZA, as in (5):

136
(4) Wo jubiqi zhe kuai shitou.  
I cannot:lift this CL stone

"I cannot lift the stone."

(5) Wo jubiqi  "I cannot:lift
A.

"I cannot lift (it)."

It seems that this is to be taken as a language-specific feature of Chinese.

Furthermore, it is obvious that Chinese displays a marked reluctance to encode inanimate referents by means of PA. As we compare the number of inanimate pronouns with the total number of pronominal anaphors as given in Table 4:5, we find that only 2 percent of all the pronouns are of inanimate reference. Following is an example:
In (6)e. the cap is encoded in PA. The scarcity of such inanimate PA in Chinese stands in strong contrast with the situation as found in English. In one investigation of English discourse (Thavenius 1983:182), it is reported that the absolute frequency of it is higher than he and she combined (1144(62%) vs. 713(38%)). Later on in the present chapter, I will discuss the implications of this phenomenon.

4.34 **Interim Summary**

The results reported so far have outlined a multi-dimensional chart for the distribution of the three anaphoric types in the data. By now, the correlations between the referential choices, on the one
hand, and discourse, syntactic, and semantic parameters, on the other hand, have been roughly established through the quantitative investigations of all the tokens found in the twenty pear stories. The following diagram summarizes the correlation between the discourse, syntactic and semantic hierarchies and the anaphoric choices:

```
Less Referential Distance/ Less Intervening Referents > More
Referential Distance/ More Intervening Referents
Pre-VBL 1 > Post-VBL 1 > Pivotal Obj > Oblique Obj > Post-VBL 2
AGT > DAT > BEN > PAT > LOC
Zero > Pronominal > Nominal
```

**DIAGRAM 4:6**

The higher up the referent is in the first three hierarchies, the more likely it is to be encoded by the type higher up in the anaphora hierarchy. On the basis of the data, the tendency is clear.

In addition to demonstrating how the two discourse parameters, distance and interference, are related to the choice of the three anaphoric types, the data also reveal how the hierarchy of anaphoric types maps onto the hierarchies of syntactic slots and semantic case roles along the dimension of continuity of the referents involved.

4.4 **Needs for Further Elaborations**

139
While the survey of the data and the discussions conducted above have provided an overview of the distribution of the three major anaphoric types in Chinese discourse, there are some important questions which remain to be addressed before we can claim a thorough understanding of the topic. Given the general correlation between the anaphoric choices and the various semantic, syntactic, and discourse factors as presented in Diagram 4:6, we have to inquire into the situations which deviate from the general tendency as has been revealed so far. For instance, cases are found in the data where ZA or PA is used to encode a referent that doesn't seem to register high continuity or predictability in discourse, as when it encodes a subject that switches reference from that of the prior sentence to another referent. Also, NA is found in situations in which the anaphor is adjacent to its antecedent and there are no competing candidates for interpretation of the anaphora. Obviously, a comprehensive treatment of the use of anaphora in Chinese discourse requires more than the rough description of the distribution of the major anaphoric types presented in the above seven tables.

With regard to the relationship between anaphoric types and topic identification, Givón (1983a) proposes a scale similar to the one given in Diagram 4:1 above. Underlying the scale, according to him, is the simple iconicity principle:
The more disruptive, surprising, discontinuous or hard to process a topic is, the more coding material must be assigned to it.

Surely this principle accounts for the distribution of the three major types of anaphoric encoding in Chinese to a certain extent, as illustrated by our discussions of the above eight tables. However, we still have two issues in front of us which have yet to be addressed. First, to understand how ZA or PA is used in spite of the apparent ambiguity that is involved with the reference of the anaphor, as can be seen from the measurements along the parameters mentioned in Tables 4:2, 4:3, and 4:4, we need a more sophisticated investigation into the ways in which the continuity of referents is assessed in addition to those used in the above tables. Second, to understand why one type of anaphoric device is predominantly preferred over the other two types when the latter will just as well fulfill the identification function, we need to investigate what determining factors other than ease or difficulty of identification underlie the choice of anaphoric types in Chinese discourse. The in-depth studies to be conducted in the next part of the chapter are aimed towards solving these two questions.

In the following several subsections, the three major types of anaphora are subject to scrutiny in turn. The survey of the twenty pear stories and the data from other sources show that more tokens of anaphora are found in the subject slot than in all other slots combined. Moreover, most of the deviations from the tendency in the
anaphoric choices as established so far are also found in this position. Taking this situation into consideration, I will concentrate on anaphora in the subject position. All tokens of anaphora will be divided into two categories, those in subject position, and those in non-subject position. Furthermore, a distinction is drawn between maintained reference and switched reference for the subject anaphors. As reported in the literature (cf. Munro, 1983), the distinction between maintained vs. switched reference in this position plays a crucial role in the anaphoric choice; hence the two situations will be treated separately.

In the following presentation, a separate subsection is devoted to each of the three anaphoric types. We will begin with the discussion of the use of ZA, which is followed by PA and NA.
4.5 **Use of Zero Anaphora**

4.5.1 **Preliminary**

ZA may encode a referent in subject position which is coreferential with the subject in the preceding clause. I will begin with discussing the use of ZA in such a situation in 4.5.2. Also, ZA may encode a referent in the subject position that switches reference from the subject referent of the preceding clause, as in the following example:

(1) a. Ta meiyou qi duo yuan, he not bike much far

b. ZA pengdao lingwai yi ge nuhaizi, come:across another one CL girl

c. ZA ye shi qi zhe zixingche. also be ride DUR bike

d. Eng, ZA guzhe kan ne ge nuhaizi...... uh concerned watch that CL girl

a. "He didn't go very far before (he) came
b. across a girl."
c. (She) was also on a bike.
d. Uh, (he) was focused on the girl..."

(S-17)
where the ZA subject in (1)c. switches reference from the referent of the subject of the previous clause to the newly-introduced girl, and the ZA subject in (1)d. switches back to the running theme that was replaced in the immediately preceding clause (1)c. Since ZA by definition does not carry any information in itself with regard to the specification of the referent it encodes, the question arises naturally what there is in the discourse that triggers the use of ZA by the speaker against this apparent ambiguity of the topic identification. This issue will be discussed in §4.43.

In §4.54, I will attempt to discuss the two major features which are identified as characteristic of the anaphors that receive ZA encoding in discourse.

4.52 ZA for Maintained Reference Subject

The most frequent use of ZA in Chinese, and perhaps in other languages as well, is to encode an anaphor that is coreferential with the subject of the preceding clause within the same sentence boundary. Consider the following examples:

\[(1)\ a. \text{Nei san ge haizi kandao ta de maozi,} \]
\[
\text{that three CL kid see he NOM cap}
\]
\[
b. \emptyset jiu cui koushao jiao ta huihai, \]
\[
\text{then blow whistle ask he return}
\]
c.  ⚫ jiu  na  gei  ta......
then  take  give  he

a. "The three kids saw his cap,
b. and then (they) whistled to call him back,
c. and then (they) give to him...."

(S-12)

(2) a. Ta suiran meiyou shuo ganxie tamen de
he  although  not  say  thank  they  NOM

hus,
word

b. ⚫ haishi hen gaoxingde qi shang le zixingche,
still  very  pleased  ride  up  PFV  bike

c. ⚫ zou  le.
leave  GRS

a. "Although he (Bike Boy) didn’t say any word of
appreciation to them,
b. (he) got on the bike,
c. and (he) left."

(S-14)

In the above examples, ZA encodes the subject that is
coreferential with the subject of the preceding clause in the same
sentence. In the great majority cases, the clauses within the sentence
boundary are subsumed under the same relational predicate. The
organization of the clauses comprising (1) and (2) is schematically
represented as (1)’ and (2)’ respectively:
It is interesting to note that in clauses subsumed under an adjoining predicate, backward anaphora, or cataphora, may also be found, which encodes a subject coreferential with the subject in the following clause within the relational predicate. In this respect the two major categories of relational predicates as discussed in Chapter Three display a marked difference.

In the adjoining predicates, the anaphor receiving ZA encoding may precede its "antecedent" instead of observing the antecedent-preceding-ZA order when the initial clause serves as the adjunct to the following nuclear clause containing the "antecedent" of the anaphora. Such "backward anaphora", or cataphora, is never found among the constituents of the conjoining predicates. Consider the following examples:
(3) a. You ye ge xiaohaizi...(4 clauses)....
   have one CL kid

   b. Zou dao lu shang de shihou,  
      walk to road up NOM time

   c. yinwei chajianerguo yi ge muhai,  
      because brush:past one CL girl

   d. ta jiu kan le ta yi yan.  
      he then see PPV he one glance

   a. "There was a kid...(4 clauses)...
   b. When (he) went up the road,
   c. since (he) brushed past a girl,
   d. he had a glance at her."

   (S-1)

   Background

   b

   Reason

   c   d

(4) a. Yinwei ji yu zuo xie chenme,  
      because eager to do some thing

   b. Ruan Min qinkan xuewen.  
      Ruan Min neglect schooling
c. 老人 qinkan xuewen,  
   because neglect schooling

   d. manman ta xiguan yu landuo.  
   gradually he accustomed to lazy

   a. "Because (he) was eager to do something, 
   b. Ruan Min neglected schooling. 
   c. Because (he) neglected schooling, 
   d. gradually he got used to laziness. (Lao She-LX)

In the above sentences, ZA is used to encode the subject of the 
adjunct clause preceding the nuclear clause with a coreferential 
subject in an anaphoric type other than ZA. The 
ZA-preceding-antecedent order, however, is not acceptable if the 
adjunct clause follows the nucleus within the relational predicate. 
Thus, (5) below can be rephrased as (6), but not as (7):
(5) a. Xiangzi bu xiang dajia,  
Xiangzi not want fight,  

b. suiran ye bu pa dajia  
although also not fear fight  

a. "Xiangzi doesn't want to fight,  
b. although (he) is not afraid of it."  
(Lao She - LX)  

(6) a. Suiran yue bu pa dajia,  
although not fear fight  

b. Xiangzi bu xiang dajia.  
Xiangzi not want fight  

Concession

149
(7) a. *Ø Bu xiang dajia,
not want fight

b. suiran Xiangzi bu pa dajia.
although Xiangzi not fear fight

Concession

On the other hand, the antecedent-preceding-ZA order as displayed in (5), which is the norm in clauses within conjoining predicates, is also found in clauses within adjoining predicates. The case study of the twenty pear stories has demonstrated that the two major groups of relational predicates don’t display any significant difference in the number of clauses across which ZA maintains reference from the prior subject. Table 4:9 gives the results of the survey:
Conjoining Predicate 1.96 clauses (n = 132)

Adjoining Predicate 2.05 clauses (n = 24)

TABLE 4:9

Average distance of ZA Maintaining Reference from Prior Subject

In the twenty-four tokens of adjoining relational predicates under investigation, the adjunct clause precedes the nucleus without exception, as exemplified in (6) above. When the adjunct clause follows the nucleus in a adjoining relational predicate, however, not one instance is found in the twenty stories in which ZA encodes a subject that maintains reference from the preceding nucleus within the same predicate. In other words, the pattern as displayed in (5) above, which, although grammatical, is rarely used in actual discourse, especially in spoken discourse.

The observations here conform to the findings reported elsewhere in the literature (cf. Li & Thompson, 1979; Chen, 1984). The cause for the phenomenon is to be understood in connection with the ordering conventions of the relational predicates. For most adjoining predicates, the nucleus-preceding-adjunct is the marked order, as
evidenced by the case with the twenty-four tokens of adjoining relational predicates in the pear stories. When the normal order between the nuclear and the adjunct clause within the adjoining predicate is shifted, two concomitant consequences follow. First, connective elements are usually needed before the adjunct clause to indicate the status of the clause; second, in spite of the role of the connective elements as indicators, the specific relation between the adjunct clause and the preceding nuclear clause is attenuated through the order shift, usually reducing the adjunct clause to a rather general "afterthought" to the nucleus. Both the effects, i.e., insertion of connectives and attenuation in rhetorical relations, are considered to be impairing to the conjoinability between the clauses involved, as evidenced by the data counts and experiments reported in Li and Thompson (1979), and Chen (1984). It is in conformity with our expectation that no instance has been found in the pear stories in which ZA encodes the subject of an adjunct clause that maintains reference from the subject of the preceding nuclear clause within the same relational predicate.

So far, we have been discussing the use of ZA for the subject that maintains reference from the coreferential subject referent of another clause within the sentence boundary. While in the Chinese pear stories ZA runs across an average of two clauses for maintained reference subject, as reported in Table 4:9, it may sometimes extend to quite a long string of clauses, separated by sentence boundaries, that are involved in a complicated relational predicate composed of
multi-levelled relational predicates. Consider the following example, in which ZA encodes the subject maintained across nine clauses in two separate sentences:

(8) a. Hu Guliang yixiang yediaowuqiang guan le,
    Hu Guliang always unconstrained used CRS

   b. jintian tou shang jiao xia dou daban zhe,
      today head up foot down all dressed:up DUR

   c. erqi de zhuangmozuoyangde yinchou keren,
      and have:to in:an:affected:way entertain guest

   d. ji wei tao dajia de chengzhan,
      both for get all NOM praise

   e. ye wei zai Xiangzi mianqian louyishour.
      also for at Xiangzi front show:off

   f. Shangbantian dao juede zhe guai youyisi.
      morning though fell this quite interesting

   g. Gang dao guowu, yin you dian pifa,
      just to afternoon because have a:little tired

   h. jiu juechu taoyan,
      then feel bored

   i. ye po xiang zhaoshen jiaoma yi chang.
      also quite want find someone quarrel one CL

a. "Hu Guliang, accustomed to free and unconstrained manners,
b. was all dressed up,
c. and had to treat the guests with affected courtesy,
d. for the purpose of inviting some praises from the others,
e. as well as showing off before Xiangzi.
f. In the morning (she) found it quite interesting, though.
g. Just past noon, (she) was a bit tired,
h. (she) began to feel bored,
i. and felt like quarrelling with someone."

   (Lao She - LX)
The propositions above share Hu Culiang as their subject. ZA is used throughout the whole passage, which depicts the referent of the anaphora in the form of a typical topic chain in Chinese.

In the twenty pear stories, as compared to 254 instances of ZA that encodes the maintained-reference subject within the sentence boundary, I have found 129 instances of ZA that encode the subject of a sentence-initial clause that maintains reference from the subject of the prior clause separated by a sentence boundary. Consider the following example:

(9) a. Tamen kan ta diedao le,
    they see he fall CRS

b. $\forall$ ba ta fu qilai.
    BA he help up

c. Ranhou 0 ye bang ta ba shuigu nong le,
    then also help he BA fruit do CRS

d. $\forall$ rang ta shang che.
    let he up bike

a. "They (Threesome) saw him fall down,
b. (they) helped him up.
c. Then, (they) also helped him with the fruit,
d. (they) let him get on the bike.
(S-16)

where (9)b. and (9)c. are in separate sentences. The subject of (9)c. receives ZA encoding in spite of the sentence boundary. Note that the ZA subject in (9)c. is preceded by a connective expression ranhou
'then'. In the 129 instances of ZA for maintained-reference subject of sentence-initial clauses, 61 (47%) of them are accompanied by expressions like ranhou, suoyi 'so', jiegou 'as a result'.

For the sentence-initial subject anaphor that maintains reference from the subject of the prior clause separated by the sentence boundary, the speaker obviously can choose between ZA or non-ZA for its encoding. As we will see later, in the majority of cases, we have PA or NA here.
4.53 ZA for Switched Reference Subject

4.531 Preliminary

It has been generally attested in various versions of the functional sentence perspective that among the syntactic slots of a clause, the subject stands the highest in the topicality (cf. Firbes 1966, Kuno 1973, Givón 1978, 1984 inter alia). When a subject is encoded in terms of ZA, people tend to take the referent of the prior subject in the preceding clause as its antecedent by default when such assumption is not semantically incompatible. It is to be expected that explicit anaphoric devices are required when the reference of the subject of the prior clause is discontinued in the current predication. In the data, however, ZA is actually found to encode the subject when it switches reference from the prior one. Following is an example:
(1) a. Na ge xiaozi hen ganxie zhe san ge ren.
that CL kid very thank this three CL person

b. Suoyi, su huan ta maozi de shihou,
therefore return he cap NOM time

c. ta jiu gei......
he then give

a. "The kid was very appreciative to the threesome,
b. Therefore, when (they) gave the cap back to him,
c. he gave...."

(S-10)

The subject of (1)b. is coreferential with the object the threesome instead of the subject in the preceding clause. It receives ZA encoding in spite of the switched reference.

I have examined all the instances where the subject anaphor switches reference from the subject of the prior clause in the twenty pear stories. The following table shows the anaphoric types of the switch reference subjects in the twenty pear stories:

157
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ZA</td>
<td>4</td>
<td>(3%)</td>
</tr>
<tr>
<td>PA</td>
<td>38</td>
<td>(25%)</td>
</tr>
<tr>
<td>NA</td>
<td>109</td>
<td>(72%)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>151</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 4:10**  
Anaphoric Types of Switched Reference Subjects in Fears Stories

In writing, ZA as switched-reference subject is more frequently used than in speaking. In the following presentation, I will also use some examples from the written data to illustrate my point. Taken in isolation, the use of ZA in such situation appears bewildering because it would be predicted to result in misunderstanding or ambiguity in the identification of the reference of the anaphor since ZA itself doesn’t carry any information with regard to the identity. In actual discourse, however, this apparently deviant anaphoric choice hardly presents any problem to the addressees, who can almost always figure out which referent in the discourse is intended by the ZA. We have two questions to address here. First, we wonder how the addressee determines that the addressee won’t take the preceding subject referent as being carried over to the ZA slot here. To put it in different terms, we are interested in what linguistic and extralinguistic factors are at work to prevent disfunctional ambiguity when ZA is used in such situations, in violation of the assumed basic
strategy of same-referent interpretation for all following subjects encoded by ZA. Second, the assurance that the addressee will correctly infer the identity of the anaphora on the basis of the linguistic and extralinguistic information does not necessarily mean that ZA must be used. If the identification of reference is the sole purpose for the use of anaphoric device, the other two types of anaphora will work just as well, if not better. We wonder what other factors are at work that contribute to the choice of ZA over PA or RA here. In our investigation of the issue, we will not confine ourselves to the use of ZA for switched reference only, but will look for underlying factors that will offer a coherent principled account for the choice of ZA for both switched and maintained reference. We will answer the first question in the next section, and allocate a separate section, §4.54, to discuss the second question.

4.532 Inference-based Identification

Since ZA itself doesn’t convey any information as to its identity, the addressee has to undergo a process of identification that is inference-based before he recognizes which referent is intended by the ZA slot here. To achieve the goal, the addressee must muster all his linguistic and extralinguistic knowledge which provides the basis for the inference process. Naturally, the addressee must make sure that such information is already available to the addressee.
for easy identification when ZA is employed for the referent. Otherwise, the use of ZA is not justified and would lead to failure in the communicative goals. It is neither practical nor necessary to explicate in the present study all the kinds of information present in discourse that contribute to the success of identification, since what is involved here is nothing less than all aspects of the linguistic and extralinguistic knowledge at the disposal of the participants in the situation, including their physical, social, cultural background, etc. Below, for illustration, we will discuss only two types of linguistic information that are provided by the discourse for the addressee in his identification process.

**Lexical Incompatibility**

First, we single out the role played by the semantics of the NPs and the predicate involved in the discourse. The selectional restrictions of the predicate may exclude the subject referent of the preceding clause as the candidate for the current subject so that the only other referent available in the context, usually the one in the object slot of the prior clause, naturally comes forward in the inferential process as the candidate antecedent of the current subject encoded as ZA. Consider the following example:
Since the Chinese word chujia used in (1)b. applies to females only, the subject Liu Siye in (1)a. is excluded as a possible candidate for the ZA slot in (1)b, leaving the other referent "his daughter" as the only candidate available in the context space for the antecedent of the ZA.

It is very common to find the antecedent of the current subject encoded in ZA occurring syntactically as a determiner in the nominal subject of the prior clause within the same relational predicate. As the referent of the prior subject is semantically incompatible with the current predicate, it is automatically rejected as the referent of the ZA subject of the predicate. Consider the following example:
Here the selectional restrictions of the lexical items involved rule out the interpretation that the left elbow could not fall asleep at midnight. The addressee can easily infer that the antecedent of the subject ZA in (2)b. is Xiangzi which turns up as the possessive in the subject NP of the prior clause, since it is the only referent available in the discourse that is logically compatible with the predicate of the second clause. Following are two more examples:

(3) a. Xiangzi de xin zhong hen luan,
    Xiangzi NOM heart in very disturbed

b. G moliao tingdao Taitai shuo pa xue,
   in:the:end hear Madam say fear blood

c. G shihu zhaodao le yi jian keyi anwei
    seem find PFV one CL can console

    ta de shi.
    she NOM thing

a. "Xiangzi’s heart was very disturbed,
b. when (he) heard the hostess talk about her fear of blood,
c. (he) seemed to have found something that could console her."

    (Lao She - LX)
   station up many people wait DUR Jiao Commissioner
b. Wen Boshi yu Tang Xiansheng de mingpian di shangqu,  
   Wen Dr. and Tang Mr. NOM card send up
   yet not wait interview
c. β hai mei dengdao chuanjian,  
   yet not wait interview
d. che you kai le.  
   train again start CRS

   a. "There were a lot of people waiting for Commissioner Jiao at the railway station.
   b. The cards of Dr. Wen and Mr. Tang were handed in.
   c. before (they) were summoned,
   d. the train left."

   (Lao She - LX)

The above three sentences, (2), (3), and (4), can be taken as illustrating the distinction between the topic and subject in Chinese. (2)a. and (3)a. have Xiangzi de you zhou ‘Xiangzi’s right elbow’ and Xiangzi’s xin zhong ‘Xiangzi’s heart’ respectively as the subject, while the topic is Xiangzi for both cases. And, the subject in (4)b. is Wen Boshi yu Tang Xiansheng de mingpian ‘Dr. Wen’s and Mr. Tang’s cards’ while the topic is Dr. Wen and Mr. Tang. The clauses that follow (2)a., (3)a., and (4)b. carry over the topic, rather than the subject, of the respective preceding clause for the referent of their subject encoded in ZA.

As we mentioned above, the most likely candidate for the ZA in the subject position is the subject of the prior clause within the same sentence. This candidate, however, can also be excluded from the interpretation for the ZA slot through inference if it is also
involved in the clause with the ZA as subject, occupying a non-subject position in the form of explicit lexical encoding that is distinctive enough for its identity. Consider the following example:

(6) a. Ta qi zixingche de shihou, he ride bike NOM time
    b. yirminian lai le lingwai yi ge nuhaizi, opposite come PFV other one CL girl
    c. $\sigma$ ye qi zhe yi liang jisodache, also ride DUR one CL bike
    d. Eng, $\sigma$ guzhe kan nei ge nuhaizi, eh concerned watch that CL girl
    e. jieguo, $\sigma$ pengdao ge pengdao yi ge shitou, as:a: result hit:against CL hit:against one CL stone
    f. $\sigma$ jiu diedao le, then fall CRS

 a. "When he was on his bike,
b. there came a girl from the other direction,
c. (she) was also on a bike.
d. Uh, (he) paid all his attention to the girl,
e. as a result, (he) hit against a stone,
f. and (he) fell down."

(S-19)

The subject of (6)d. switches from the subject of (6)c. back to the one that was introduced in (6)a., but was displaced in (6)b. and (6)c. The morphological encoding of the subject here, however, doesn’t signal the shift in reference in an obvious way. For the subject of (6)d. encoded as ZA, there are two competing candidates for its identity, the girl as the ongoing subject from (6)c., and the
displaced subject from (6)a. The appearance of the girl in NA encoding in the object position of (6)d., however, disqualifies her as the subject of the clause, and leaves the displaced subject the bike boy as the only eligible antecedent in the context for the ZA subject of (6)d.
Syntactic Parallelism

Another important source of disambiguating information as part of the linguistic knowledge of the addressee is provided by the syntactic configurations of the clauses that contain the anaphors and relevant referents in the discourse. In this respect, we have found in the parallelism of syntactic configurations one of the most important factors contributing to the identification of the referent encoded in ZA as the switched-reference subject.\(^3\)

Syntactic parallelism is a specific kind of conjoining relationship characterized by identity or similarity of the syntactic configurations of the constituents, often accompanied by a similarity at other levels such as lexical, semantic, or rhetorical. It is observed at low levels, such as lexical, and phrasal, as well as at higher levels, such as clausal, or rhetorical. Consider the following example:

(7) a. Ta qu,
    he go

    b. wo jiu qu;
    I then go
c. ta bu qu,
   he not go

d. wo ye bu qu.
   I also not go

a. "If he goes,
b. I will also go;
c. if he does not go,
d. I won't go, either."

The conjoined parts of the top-level relational predicate display a parallelism between themselves: a marked similarity in terms of syntactic compositions as well as lexical items and inter-clausal relations between (7)a. and (7)b., on the one hand, and (7)c. and (7)d., on the other.

It is precisely such similarity in the patterning of the constituent parts, compounded with other linguistic and extralinguistic information, that provides the clues to the successful inference of the intended referent for ZA encoding the switched-reference subject.
Parallelism in patterning contributes to the identification of referents at various positions in discourse, including those as switched-reference subject both in adjunct and nuclear clauses in the relational predicates. Following is an example illustrating ZA for the referent as switched-reference subject in the adjunct clause:

(8) a. Ta zhi neng you bieren de hua zhong
    he only can from others NOM word in
    xishou xie shengming de kuwei.
    absorb some life NOM bitterness

b. Daji dou kunao,
    everyone all worry

c. ta ye bu shi liwai.
    he also not be exception

d. Daji shou dao beiku de difang,
    everyone say to sad NOM place

e. ta ye zhoushang mei;
    he also knit eyebrow

f. shuo dao kexiao de difang,
    say to funny NOM place

g. ta ye piepie zul.
    he also twitch mouth

```
  a. "He can only have some taste of bitterness of life from
     the others' words.
  b. When the others were all worried,
  c. he was no exception.
  d. When they talked about something miserable,
  e. he also had knit eyebrows;
  f. when (they) talked about something funny,
  g. he gave some grins, too."
```

(Lao She - LX)
In the above string of clauses, (8)f. and (8)g. follow the same pattern as (8)d. and (8)e. in terms of lexical, syntactical, and rhetorical composition: The lexical expressions are the same or symmetrical; the syntactic structure is the same; the clauses bear the same rhetorical relation between them. On the basis of the parallelism, the referent of the ZA subject of (8)f. can easily be inferred as coreferential with the referent in the corresponding position in (8)d.

In (8), the clause with the ZA subject and the corresponding clause are both adjuncts in their respective relational predicate. Note that ZA as switched reference subject may also appear in a nuclear clause when the parallelism at issue is present. Consider the following example:
(9) a. Li Li jiao le Zhishu,  
Li Li call PFV secretary  
b. Zhishu bing bu na dao;  
secretary but not take knife  
c. Shijia jiao le Duizhang,  
call PFV team:leader  
d. Duizhang ye bu na dao.  
team:leader also not take knife  
e. Dajia yiqi shang shan.  
all together go:up mountain  
a. "Li Li called on the Secretary,  
b. but the Secretary didn’t carry any knife;  
c. (Li Li) called on the Team Director,  
d. the Team Director didn’t carry any knife, either.  
e. They went up the mountain together." (Zhong - QW)  

In the above example, the identity of ZA as the switched reference subject in the nuclear clause (9)c. can be inferred on the basis of the parallelism of the lexical expressions, syntactic configurations, and rhetorical relations between (9)a. and (9)b., and (9)c. and (9)d.
Furthermore, parallelism in the corresponding constituents of the relational predicate can lead to the identification of ZA as the switched reference subject in both adjunct and nuclear clauses, as shown by the following example:

(10) a. Ta yikouqi pao dao beikou, he straight run to north:gate
b. zixingche hai gen zhe ne! bike still follow DUR SPL
c. Ta jin le xiao hutong, he enter PFV small alley
d. Zhai hai gen zhe. still follow DUR
e. Zhai le hutong, go:out PFV alley
f. Zhai hai gen zhe. still follow DUR

a. "He ran to South Gate without stop,
b. The bike was still following.
c. He entered a small alley,
d. (it) was still following.
e. (He) went out the alley,
f. (it) was still following."

(Lao She - LX)
The six clauses above fall into three pairs, (10)a. and (10)b., (10)c. and (10)d., and (10)e. and (10)f. Parallelism in lexical, syntactic, and rhetorical organizations is present among the three pairs. Within each pair, the first clause provides the background information for the second, indicating the spatial environment in which the bike followed the man. The identification of the subject in (10)d., which receives ZA encoding in spite of the switched reference, is based upon the subject of the corresponding clause (10)b. in the first pair. The same principle applies to the identification of the ZA subject of the two clauses in the third pair.

**Interim Summary**

In this section, we have discussed the two factors that invalidate the default assumption that the subject encoded in ZA generally carries over reference from the prior subject, thus accounting for the use of ZA as switched reference subject. We must add that, in the final analysis, the two factors that we have chosen to illustrate above, lexical incompatibility and parallelism, are only two out of all the aspects of linguistic and extralinguistic information that are resorted to in the resolution of ZA as switched reference subject. The progress of science in this domain depends upon our gradual in-depth investigation of the characteristics of all the factors involved in the identification process and the interrelationships thereof.
In the discussion of the use of PA, we will show that there are cases in which the switched-reference subject may receive PA encoding when the pronoun itself cannot differentiate the anaphor from the competing referents in the discourse. What has been said in the identification of ZA that encodes the switched-reference subject also applies to PA in the same situation.
4.54 Explaining Choice of ZA

4.541 Preliminary

So far we have discussed how the knowledge provided by the context has contributed to the unambiguous identification of the referent encoded in ZA as switched reference subject. However, as we mentioned earlier, the fact that the identity of the referent can be inferred on the basis of linguistic and extralinguistic knowledge does not necessarily mean that it must receive ZA encoding. In many situations, PA or NA is chosen in spite of the fact that no potential ambiguity in the anaphora identification would result if ZA had been used for the referent. In other words, being subject to correct identification through inference stands as a necessary condition for triggering ZA, but not as a sufficient condition. What we are interested in here is the question of what motivates the preference of ZA over PA or NA in such situations.

A careful study of all the relevant tokens in our data has led to identifying two features which are correlated with the selection of ZA as maintained reference subject, or as switched reference subject when the identity of the anaphor can be inferred on the basis of the linguistic and extralinguistic knowledge of the discourse.
participants. Both of the features are discourse related. One is high continuity of the referent. The other is high negligibility of the referent. Next, we will detail upon each of them.

4.542 High Continuity of Referent

As Chafe (1979) proposes, continuity of the referent is one of the major aspects in which phrases achieve coherence in discourse, the other being continuity of space, time, event, and world.

Continuity of the referent relates to how easily the referent can be identified in the discourse. As discussed in Givón (1983b), it is a complex functional domain that constitutes a scalar, gradual continuum that ranges from the most continuous, or most easily identified, to the least continuous, or least easily identified.

Previous studies have amply demonstrated that the position of referents along the continuity continuum in the discourse is intimately tied with a variety of linguistic phenomena in various languages, including anaphoric choices in various languages (cf. studies collected in Givón 1983a) and particle movement in English (cf. Chen 1986). Linguists disagree, however, on the issue of how the continuity of referents is established in discourse. The following discourse measurements are adopted to assess the topic continuity of referents in Givón (1983b), and also in Chen (1986) with slight modifications:

175
1. **Referential Distance**, which assesses the gap between the previous occurrence in the discourse of a referent and its current occurrence in a clause, where it is marked with a particular grammatical coding device.

2. **Potential Interference**, which assesses the interference of other referents within the immediate preceding register that may compete for the identity of the referent in issue. This measurement is established in terms of the number of such potentially competing referents within the five immediately preceding clauses.

3. **Persistence**, which reflects the importance of the topic in the discourse and measure the speaker’s topical intent. It is assessed in terms of the number of mention in subsequent discussion in which it continuous as an uninterrupted presence as a semantic argument of the clause, an argument of whatever role and marked by whatever grammatical means.

The three measurements have also been adopted in the present thesis in the preparation of Tables 4:2, 4:3, 4:4, and 4:5. The results as displayed in Tables 4:2, 4:3, and 4:4 have attested to a general correlation between the scales of the continuum and the encoding devices in Chinese as represented in Diagram 4:1. Roughly speaking, ZA
is more likely to be chosen when the anaphor is very close to its antecedent in the linear order and there is no interfering referent between the two.

Out of the total instances of ZA in subject position as found in the pear stories, 81% (383:473) maintain reference from the subject of the prior clause, and 15% (69:473) maintains reference from the referent in postverbal position of the prior clause. Only 4 instances are found in which ZA encodes the subject anaphor that switches reference from the prior subject to its antecedent that appears earlier in discourse. The above distribution indicates that high continuity of reference is a necessary condition for triggering ZA in Chinese.

However, it doesn't follow from the above discussion that to be high in continuity of reference necessarily triggers ZA. We will see later in the study of PA and NA that the other two anaphoric types are also used for subject anaphors that maintain reference from the prior clause. Among the 452 instances of ZA encoding subject that maintains reference from the prior subject or the postverbal argument of the prior predicate, we have established that 89% (402) of them are united with the prior clause through involvement with a single schema and/or involvement with a single goal. Consider the following example:
(1) a. Duimian lai le yi ge xiao de nuhaizi,
   opposite come PFV one CL small NOM girl

   b. ye shi qi, ... qi jiaodahe de.
      also be ride ride bike NOM

   a. "From opposite came a small girl,
   b. (she) was also on a bike."
      (S-13)

where (1)a. and (1)b. focus on the categorization of the character and
the action she was engaged in. The two clauses are integrated through
involvement with what Chafe (1979) calls character-action-(location)
schema. Following is another example illustrating clauses united
through involvement with a single goal:

(2) a. San ge nanhai jiu pao guolai bang ta,
      three CL boy then run over help he

   b. ba ta fu qilai,
      BA he help up

   c. ba ne ge di shang de shuiguodou
      BA that CL ground up NOM fruit all

      gei jian qilai.
      give pick up

   a. "The three boys ran over to help him,
   b. (they) helped him up,
   c. and picked up the fruit on the ground."
      (S-20)

where the actions in (2)a., (2)b., and (2)c. constitute a coherent
chain of events that leads to the realization of a single goal --
helping the fallen bike boy.
The above discussion demonstrates that ZA is more frequently used when the continuity of reference is coupled with other integrating principles of discourse such as involvement with a single schema and/or with a single goal as the clause that contains its antecedent.

In §4.53, moreover, we have shown that ZA may also be used when the continuity of the referent is broken, as assessed in terms of the three measurements based on the linear order, as in the following example:

(3) a. Ta (Bike Boy) qi, qi, qi dao yiban...(2 clauses)...
   he ride ride ride to half

   b. Qianmian lai le yi ge nuhaizi,
      front come PPV one CL girl

   c. $\&$ ye qi she zixingche.
      also ride DUR bike

   d. $\&$ Guzhe kan na ge nuhaizi......
      concerned:with watch that CL girl

   a. "He was riding half down the way...(2 clauses)
   b. Opposite came a girl,
   c. (she) was also on a bike,
   d. (He) was too occupied with watching the girl...."
      (5-11)

where ZA is used in (3)d. in spite of the fact that there is a gap of two clauses, (3)b. and (3)c. with switched reference subject, between the anaphor and its antecedent that appears earlier in the discourse.

Scrutiny of cases like (3) within the hierarchical structure of discourse has led many linguists to the conclusion that the position
of the referent in the hierarchical structure of the discourse is a more revealing way to establish the continuity of the referent.

As mentioned earlier in Chapter 3, the position of a referent in discourse can be located from two perspectives, the linear and the hierarchical. I maintain that the anaphoric choice of the referent is sensitive to its position in both the linear and the hierarchical organization of the discourse. The distance and interference theory, which is based upon the linear order of the referent, is not completely falsified by cases like (3). Instead, these cases should be interpreted as indicating that the application of the theory should be sensitive to hierarchical structure. It holds when the anaphor, the antecedent, and the possible interfering referents are on the same level in the hierarchical structure of the discourse. The intervening elements are much less disruptive to the continuity of the referent if they are on different levels. When a anaphor pops over the linear intervening referents at a different level back to its antecedent, it may still be high in continuity if the antecedent has been the running theme in the discourse, and if the popped-over material is not very heavy in composition. In such a situation, the hierarchical closeness at the same hierarchical level overrides the linear distance and intervention. Consider the following example that illustrates this point.
(4) a. Zhang Dashao zuocai,
    Zhang Dashao cook

    b. ₃ duan cha,
        serve tea

    c. ₃ rang keren,
        entertain guest

    d. ₃ tian tang,
        add soup

    e. ₃ huan kuaizi ---
        change chopsticks

    f. Lao Li chi gao:le:xin,
        Lao Li eat pleased

    g. ₃ ba kuaizi diao zai dishang liang hui ---
        BA chopsticks drop on ground two time

    h. ₃ ziji tiao fei de chi,
        self pick fat NOM eat

    i. ₃ kuajiang ziji de shouyi,
        praise self NOM skills

    j. ₃ tongshi bing ju.
        same:time together do

a. "Zhang Dashao prepared dishes,
b. (she) served tea,
c. (she) entertained the guest,
d. (she) added soup,
e. (she) changed chopsticks---
f. Lao li was so pleased with the food that
g. (he) dropped the chopsticks twice---
h. (she) picked the fat for herself,
i. (she) praised her own cooking,
j. and (she) did all these at the same time."

    (Lao She - LH)
Note that in the above clauses, the subject in (4)f shifts from Zhang Dasao to Lao Li. From (4)h onward, however, the subject referent pops back to Zhang Dashao, which is encoded in terms of ZA here. If we examine the rhetorical structures of this segment, we find that (4)f and (4)g with intervening subject referent are not on the same hierarchical level with the other clauses. While all other clauses are conjoined under the succession predicate, Clauses (4)f. and (4)g. serve as the elaboration adjunct to Clause (4)e. Clause (4)h. with ZA for its switched reference subject pops back to the topic of its sister-node clauses for the reference of its subject across the intervening referent Lao Li. Here the linear distance and interference are overridden by the closeness of the anaphor and its antecedent on the same level in the hierarchy. Neither the gap between the anaphor and its antecedent nor the presence of a semantically compatible referent in the immediately preceding clauses prevents the use of ZA in (2)h. It is usually the case that the clauses with the intervening
topic serve as adjuncts to the clauses that contain either the antecedent of the anaphor or the anaphor itself. Following are two more examples. First let's consider this:

(5) a. Xiangxiang dao wusuowei.
    Xiangxiang but not:care

b. 'Zijia you hao chi de,
    self have good eat NOM

c. ' jiu dao zai Sanda de wan li.
    then pour in Sanda NOM bowl in

d. Sanda you hao chi de,
    Sanda have good eat NOM

e. ' ye dongshou na wan qu cheng.
    also act take bowl go fill

a. "Xiangxiang didn't care.
b. When (she) herself had some good food,
c. (she) would dish out to Sanda.
d. When Sanda had good food,
e. (she) would fill her bowl with it."

(People's Daily - 1985)
in which (5)c. and (5)e. are the respective nuclei of two separate relational predicates of Condition which are connected by a joint predicate. The subject referent of (5)e. pops over the the intervening (5)d. back to the subject of (5)c. ZA is used for the subject of (5)e. in spite of the distance between the anaphor and its antecedent, and the discontinued reference of the subject of (5)d., as a result of the fact that the popped-over clause (5)d. is not coordinately joined to the either (5)c. or (5)e., and it is not very complicated in composition. Now we present another example:

(6) a. Ta yao pai jishuyuan chuqu xuexi, he want send technician out learn
   b. Changzhang bu diantou, director not agree
   c. $zoubuliao, cannot:leave
   d. $zhao wo zhe ge fu zong gongchengshi, come:to I this CL deputy chief engineer
   e. wo ye wunengweili, I also helpless
   f. $ji de disalie, worry CSC weep

a. "He wanted to send a/some technician/s out for training,
b. However, the Director wouldn’t give permission,
c. so (he/they) couldn’t go.
d. (He) came to me, the Deputy Chief Engineer for help,
e. but I couldn’t do anything about it,
f. (he) was worried to tears."  (People’s Daily - 84)
(6) is a particularly revealing example illustrating how topic continuity is maintained in discourse. So far as the linear order is concerned, the ZA subject in (6)d. is separated from the antecedent by two intervening clauses, (6)b. and (6)c., and there are two referents in the immediately preceding register which might have a disruptive effect upon the identification of the referent in issue. The subject in (6)f. is also a pop-back to the subject in (6)d. across another referent that is semantically compatible with the predicate in (6)f. From the hierarchical perspective, it is clear that these clauses are subsumed at the top level by a Succession Predicate, which has three nuclear clauses, (6)a., (6)d., and (6)f., narrating how he wanted to do something, and went to the chief engineer for help, and was worried to tears in the end. The other clauses all serve as adjuncts to the three nuclei in the hierarchy. Despite the apparent low continuity as assessed by the linear-order-based measurements in Givon (1983), ZA is
used in (6)d. and (6)f., as a result of closeness of the two clauses to (6)a. at the top level of the hierarchy.
From the above discussion, we realize that the two linear-order-based factors, distance and interference, hold only when the intervening clauses are co-ordinately joined to the clauses containing the anaphor and its antecedent in the hierarchical organization of discourse. If the intervening clauses are adjuncts to the nuclear clauses in which the anaphor or its antecedent occurs, and are not heavy in composition, the anaphor that pops over the intervention back to its antecedent may still be high enough in continuity to receive ZA encoding.

As we have discussed above, distance and interference can serve to assess the continuity of the referent to the preceding discourse. In Table 4:5, we have also examined the continuity of the referent in three major anaphoric types to the following discourse in terms of the number of subsequent clauses in which the referent maintains an uninterrupted semantic presence. Studies in various languages in Givon (1983) have indicated a correlation between longer persistence and more attenuated anaphoric forms. It is reported that among the three major anaphoric types, referents for which ZA is used have the longest persistence, while those for which NA is used have the shortest.

Table 4:5, however, shows that Chinese differs from other languages in that it is the PA that persists the longest, with ZA coming a very close second. As a result, we cannot take high
persistence of the referent in the following discourse as a feature that contributes specifically to the choice of ZA over PA.

In summary, the above investigations have led us to the conclusion that the high continuity of the referent to its antecedent in the preceding discourse as displayed through the linear closeness at the same hierarchical level of discourse coupled with other integrating principles such as involvement with a single schema and/or with a single goal with the clause containing its antecedent is a contributing factor to the choice of ZA.

4.543 High Negligibility of Referent

In the data, however, we have found quite a few tokens of ZA the use of which cannot be explained in terms of high continuity of the referents. These instances of ZA typically encode the referents which are low in continuity to the preceding discourse by whatever means of assessment, but the identity of which can be inferred through other linguistic or extralinguistic information available in the discourse. An examination of the characteristic features of the referents involved in this situation leads us to the identification of high negligibility of the referent as the second factor (the first being high continuity of the referent) that contributes to the choice of ZA in discourse on the presumption that the identity of the anaphor can be inferred without ambiguity.
There are two major kinds of evidence which strongly suggest correlation between the high negligibility of the referent and the ZA encoding it receives.

The first kind of evidence comes from the difference in anaphoric encoding between referents of high saliency and those of low saliency in our data. First, let's consider the inherent saliency of the referents. It is generally agreed that animates are higher than inanimates along the continuum of saliency (cf. Comrie 1981). Table 4:8 demonstrates that in the twenty pear stories there are only ten instances of PA that are inanimate, constituting 2% of all PA instances. All the other occurrences of the inanimates are encoded in ZA or NA. ZA is usually chosen when the reference can be inferred from whatever information is available in the discourse; otherwise, NA is used. Consider the following example:

(1) a. Zai guolai,
    again come:along

   b. yi ge xiao haizi qi danche chuxian le.
      one CL kid ride bike appear CRS

      ... ... (13 clauses) ... ...

189
The bike, which is introduced in (1)b., is encoded in NA in (1)c., (1)d., and (1)e. It receives ZA encoding in (1)g. and (1)h., where it is easily inferred to be the missing argument of the predicate verbs.
qi 'ride', and tui 'push'. Here it is impossible to attribute the use of ZA to the high continuity of the referent for the obvious reason that there is no high continuity of reference either from the linear perspective or from the hierarchical perspective.

Situations in which ZA is chosen in spite of the low continuity of the referent so long as its identity can be inferred from the discourse and it is on the animacy hierarchy are also found with referents with low saliency. As mentioned in §2.43 and §2.44, the encodings that are accorded to the referents when they are introduced into discourse are correlated with the saliency assumed by the addressee for the referent. The referent with high saliency tends to receive determinate encodings while the one that is low in saliency tends to be lexically encoded in indeterminate terms, among which the bare noun is by far the most frequently used. I have counted all the anaphoric mentions of the sixty-one inanimate objects in the twenty pear stories that receive indeterminate encodings when introduced into discourse. They are mentioned for a total of one hundred and sixty-nine times later in the discourse. The following table shows how they are distributed among the three anaphoric types:
ZA  53  
   (31.4%) 

PA  1  
   (0.6%) 

NA  115  
   (68%) 

Total:  169  

TABLE  4:11  

Anaphoric Distribution of Indeterminate Inanimates

This shows that we have fifty-three instances of ZA, and one hundred and fifteen instances of NA. PA is used only once to encode the referents in low saliency. When PA is used to encode such referents in English; Chinese usually resort to ZA or NA. The bike in (1) is a good example illustrating how an inanimate object introduced in indeterminate terms is encoded in anaphoric terms later in the discourse. Following is an example illustrating the anaphoric encoding of a human referent introduced into the discourse in indeterminate terms. Now pay attention to gongzuoren 'functionary' in the following clauses:
(2) a. Yīcì, Mào Zèdōng Tōngzhī bìng zhòng le,
    once Mào Zèdōng comrade illness serious CRS

b. bù néng kānshū le.
    not can read CRS

c. Tā jiù qīng gōngzuòyuán dūshū gěi tā tīng.
    he then ask functionary read for he listen

   a. "Once, when Comrade Mao Zedong was seriously ill,
   b. (he) couldn’t read.
   c. So he asked the functionaries to read for him."
   (People’s Daily - 85)

As we have noticed before, a feature that is characteristic of such
referents of low plot saliency is that usually no PA is used to encode
them when they continue as a semantic argument of the following
clauses in the discourse. Now, consider the continuous occurrence of
the functionaries in (2)c. in the following discourse given below:

cont.
(2) d. Tā cèng yòng chàndòu de shǒu zài yī bēn
    he once use tremble NOM hand on one CL

   << Lu Xùn Quán Jí >> fēngmiàn shāng xiè xià
   Lu Xùn complete works cover up write down

   “chǐ lǎn píngguǒ” jí gè zi,
   eat rotten apple several CL character

e. bù shíyì gōngzuòyuán du <<Guànyǔ Fānyì
   hint functionary read about translation

   (Xīa) >> yī wén.
   last:part one article

   d. "With trembling hands, he once wrote down the characters
      'eat rotten apples' on the cover of the Complete Works
      of Lu Xun,
   e. and (he) hinted to the functionaries to read the
      article 'On Translation (Last Part)' for him."

193
When the identity of such referents can be inferred, ZA is normally used for its encoding, as exemplified by Clause (2)f. that follows (2)e. above:

(2) f. Dang du dao Lu Xun yong "wan lan pingguo" de biyu......,
    when read to Lu Xun use 'scoop:out rotten apple' NOM metaphor

  g. Mao Zedong Tongzhi liansheng jiaodao.......  
     Mao Zedong Comrade repeatedly shout

As we can see, gongzuoren yuanc ‘functionary’ is introduced in indeterminate terms into the discourse in (2)c. It receives its first anaphoric mention in the pivotal slot in (2)e. Since ZA is not allowed in such a slot, the referent is encoded in NA here. It is mentioned again in (2)f. as the subject of the clause, which switches reference from the prior subject Comrade Mao Zedong. In spite of the drop in continuity resulting from the switched reference, the anaphor receives ZA. This again shows that referents low in inherent or plot saliency usually receive ZA encoding if their identity can be easily established through linguistic and extralinguistic information available in the discourse even when they register low in continuity to the preceding discourse.

194
The other piece of evidence supporting low saliency as one contributing factor for the choice of ZA comes to light when the referent encoded in ZA is significantly low in topic continuity, as when it occurs paragraph-initially. As reported repeatedly in Tai (1978), Hinds (1981), Givón (1981), inter alia, paragraph boundaries usually mark major breaks in thematic continuity, which typically also implicate breaks in action or topic continuity. As a result, referents at paragraph-initial positions are unarguably considered to be low in continuity to the preceding discourse. Obviously, if referents in such positions are found to receive ZA, we cannot resort to high continuity as the contributing factor, and must look for other factors which are responsible for the anaphoric choice here.

I have chosen to use written materials as my data for examination, since they provide explicit paragraph boundaries. I have examined ten pages in succession taken out of the middle of Lao She's Luotuo Xiangzi (LX) with special attention to the anaphoric types of referents as found at the beginning of paragraphs. The ten-page excerpt narrates how Xiangzi fled from the mountain into the town. The examination of the excerpt has revealed that there are indeed referents encoded in ZA which occur precisely at paragraph-initial positions. Out of forty-eight paragraphs contained within the excerpt, eight have a paragraph-initial referent encoded in ZA, as exemplified in the following sentence:

195
(1) a. Shuo dao zheli,
speak to here

b. Xiangzi buyoude ьио le qilai.
Xiangzi cannot:help laugh PFV up

a. "When (he) spoke of this,

b. Xiangzi couldn't help laughing."

(Lao She - LX)

where the paragraph-initial referent as the subject in (1)s. receives ZA encoding in spite of the fact that it is in a position of low continuity with the preceding discourse.

Taking the rhetorical structure of the clauses involved into consideration, we have found that the eight tokens of paragraph-initial ZA all occur in clauses which are adjuncts to the following nuclear clauses subsumed under one of the adjoining predicates, as schematically indicated in the following diagram:

```
          /
         /
Adjunct——— Nucleus
          /
         /
ZA        PA/NA
```

DIAGRAM 4:7
In total, eleven adjoining predicates are found at the beginning of paragraphs, with eight assuming the pattern in Diagram 4:7, two assuming that in Diagram 4:8, and one assuming that in Diagram 4:9:

![Diagram 4:8]

As I have remarked earlier, it is impossible to attribute the use of ZA for paragraph-initial referents as schemed in Diagram 4:7 to the high continuity of the referent because it is well recognized that paragraph boundaries mark major breaks in discourse continuity.
Moreover, there are two additional pieces of evidence against the theory of high continuity as the only factor responsible for the use of ZA here.

First, out of the eight paragraph-initial ZA subjects in the pattern of Diagram 4:7, four have switched reference from the subject of the clause ending the previous paragraph. The average distance between the ZA and its antecedent is 3.75 clauses. As we can recognize from the previous discussions, such distance and interference from other competing referents militate strongly against the continuity of the anaphor.

Second, we have cases in which the ZA subject in the pattern of Diagram 4:7 is not only paragraph-initial, but also discourse-initial, such as at the very beginning of a story. Consider the following sentence:
Clause (2)a. occurs right at the beginning of the story. It is utterly impossible to attribute the use of ZA for the subject referent to the continuity of the referent to the preceding discourse. In fact, there is no way at all to infer the identity of the anaphor if it is interpreted apart from the following (2)b. We simply cannot explain why ZA is chosen for the subject in (2)a., and NA for the subject in (2)b., instead of the other way around, if high continuity of the referent is the only factor responsible for ZA.

In the above discussion, we have examined two situations in which it is inappropriate or impossible to attribute the use of ZA to high continuity of the referent. In the first situation, the referents in ZA are low in inherent or plot saliency, as can be revealed from the lexical encodings that they receive when introduced into discourse. In the second situation, the referents in ZA all occur in the subject slot of the clauses that are adjoined to the following nuclear clauses.
in the same relational predicate. I maintain that the feature that is characteristic of the referent receiving ZA in the two situations is their high negligibility, which is the other factor that accounts for the choice of ZA in Chinese discourse.

The negligibility factor reflects the psychological weight of the referent in the discourse. Some referents are less noteworthy than others by their nature or by their use in discourse. For instance, generally speaking, inanimates are less noteworthy than animates, and adjunct clauses, together with their constituent parts, are usually less important than nuclear clauses for the communicative goal of the interlocutors. The relationship between the noteworthiness of information and its linguistic expression can be captured in the following principle:

The less noteworthy a piece of information is,
the more attenuated is its linguistic encoding.

This principle is in conformity with the principle of economy and the principle of efficiency in verbal communications. Within the limited capacities of the short-term memory encoding and decoding devices of the interlocutors, more attention is directed to the more noteworthy information when the less noteworthy information receives encoding no heavier than necessary. So far as a referent high in negligibility in discourse is concerned, ZA is the most appropriate anaphoric device whenever its identity can be inferred from the text by whatever means.
There is no need to "highlight" it in heavier terms than what is necessary for its identification. Such a functional consideration can be grammaticalized in the use of unstressed pronouns or clitics. Also, the functional principle is grammaticalized in languages like English in the infinitival, participial, and gerundial constructions, which have the subject encoded in ZA if it is coreferential with the subject in the dominating clause. In Chinese, however, the functional principle is mainly displayed at the discourse level, underlying the choice of ZA for the referents of low inherent or plot saliency, or for referents in the subject slot of the adjunct clause which are coreferential with the subject in the following nuclear clause.

4.55 Summary

In summary, ZA may be chosen to encode a referent in discourse, provided that its identification can be established through whatever linguistic and extralinguistic information is available in discourse. The actual choice of ZA over PA or NA is attributed to one of the following two discourse-pragmatic factors: one is high continuity of the referent coupled with other discourse-integrating principles such as involvement with a single schema and/or with a single goal; the other is low noteworthiness, or negligibility, of the referent.
4.6 Use of Pronominal Anaphora

4.6.1 Preliminary

The Chinese pronominal system distinguishes singular and plural referents, but not gender either in the singular or in the plural. Thus, as compared to ZA, PA itself conveys more information in Chinese with regard to the identification of the referents, although it is less informative in this respect than in the languages characterized by a richer gender system such as French, German or English. In this section, first, we will concern ourselves with the use of PA that encodes a subject that maintains reference from the subject of the prior clause. After that, we will discuss the use of PA for a switched reference subject, where we will draw a distinction between PA that can differentiate the anaphor from a competing referent, and PA that cannot. Following will be a brief discussion of the use of PA elsewhere in the clause. In the end, we will try to answer the question of what discourse features the triggering of PA in Chinese can be attributed to.
4.62 PA for Maintained Reference Subject

When the subject referent of the present clause is maintained from the preceding clause, it doesn't mean that ZA is always the encoding type that chosen. In our data, PA or NA is also found in the position. For the present, let's consider the instances of PA which encode the subject referent that maintains reference from the prior subject.

As in the case of ZA, the use of PA here must be investigated in connection with the discourse features of the environments in which it is used, specifically in connection with the continuity of the discourse in which it occurs. As discussed earlier, discourse continuity is maintained in terms of coherence of space, time, topic, event, and so on. In the linguistic deployment of the flow of consciousness, the speaker tends to use the sentence to encode a group of phrases, typically case frames with predicate and arguments, that achieve coherence among themselves through a few principles. These principles are involvement with a single referent, with a single scheme, or with a single goal.

In the twenty pear stories, I have found 144 instances of PA that encodes the subject of a clause that maintains reference from the prior subject. Out of these 144, 126 instances are used as the subject at the beginning of a sentence. Consider the following example:
(1) a. You ge xiaohai, exist CL kid
b. qi zhe yi liang ... jiaodache, ride DUR one CL bike
c. jingguo zhexie shuigu. pass these fruit
d. Ranhou, ta ting xialai, then he stop down
e. ba yi lou shuiguo tongtong ban dao che shang. BA one CL fruit all move to bike up

a. "There was a kid,
b. (who) was riding a bike,
c. passed the fruit.
d. Then, he stopped,
e. and moved the whole basket of fruit up to the bike."

(5-20)

The subject in (1)d. is coreferential with that in (1)c. In spite of the short distance and the lack of interfering referent between the anaphor and its antecedent, the anaphor receives PA encoding.

The 126 subject referents in PA, though coreferential with the subject in the prior clause, are separated from the preceding discourse by sentence boundaries, which are marked by an intonational contour in speaking, and a final punctuation mark in writing. Obviously, the coherence between the clause in which the PA occurs and the clause containing its antecedent is disrupted here not by switched reference, but by other factors.
In the "flow model" of the linguistic deployment of thought proposed by Chafe (1979), phrases which are organized in the form of sentences are usually those that achieve unity among themselves through involvement with a single referent, with a single schema, or with a single goal. I have examined the 126 sentences which contain the sentence-initial subject PA that maintains reference from the subject in the preceding sentence to see how the sentence boundaries are drawn through involvement with different schema and/or different goal. The results are presented in Table 4:12:

<table>
<thead>
<tr>
<th>Different Schema/</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different Goal</td>
<td>(88%)</td>
</tr>
<tr>
<td>Different Schema/</td>
<td>8</td>
</tr>
<tr>
<td>Same Goal</td>
<td>(6%)</td>
</tr>
<tr>
<td>Same Schema/</td>
<td>4</td>
</tr>
<tr>
<td>Different Goal</td>
<td>(3%)</td>
</tr>
<tr>
<td>Same Scheme/</td>
<td>3</td>
</tr>
<tr>
<td>Same Goal</td>
<td>(3%)</td>
</tr>
</tbody>
</table>

**TABLE 4:12**

Features Marking Sentence Boundary Between Sentence-initial Subject PA and its Antecedent in Prior Clause

As the table shows, in the great majority of the cases (88%), the clause with a PA subject is separated from the prior clause with the coreferential subject through both different schemata and different goals. In the 126 sentences, 78 are preceded by connective expressions.
like ranhou 'then', houai 'afterwards', keshi 'however', or preceded by signs of hesitation like eng 'uh', and so on. Following is an example:

(2) a. Ne san ge nanhaizi ... paipai ta.
    that three CL boy pat him

b. Ranhou, jiu ... eng, ranhou, tamen guoqu yihou,...
    then then uh then they pass after

a. "The three boys ... gave him a pat.

b. Then, then ... uh, then, after they passed...

(S-16)

Such insertions as ranhou 'then', eng 'uh' between sentence boundaries mark segmentation in the continuous flow of discourse.

There are seven instances in which the sentence with an initial PA subject is involved with the same schema as the preceding sentence containing the antecedent of the PA. These instances fall into two broad situations. In one situation, the following sentence corrects, or modifies the content in the prior sentence, as exemplified by the following example:
(3) a. Ta (Bike Boy) jiu qi zou le.
   he      then ride go CRS

   b. Ta [dou] meiyou ... meiyou qi zou,
      he       not not ride go

   c. ∅ jiushi yiguaiyiguaide jiu zoudiao le.
      then limping then go:away CRS

   a. "He then biked away.
   b. He didn't bike away,
   c. (he) then limped away."
   (S-2)

where (3)a. in the first sentence and (3)b. and (3)c. in the second sentence are involved in the same schema in that they both narrate what the bike boy did after the threesome helped him up with the bike and pears. In this case, (3)b. and (3)c. constitute a correction to the preceding sentence. In the other situation, the second sentence repeats what has been said in the first sentence, as exemplified by the following sentences:

(4) a. Ranhou ne, ta jiu shang le ma.
   then REx he then wound CRS PAR

   b. Ta shou shang le.
      he get wound CRS

   a. "Then, he was wounded.
   b. He got wounded."
   (S-13)

where the second sentence repeats what is said in the first sentence, though not word for word. In spite of the coherence of the two
sentences in (3) and (4) through involvement with the same schema and referent, PA has been chosen here for the subject referent preceded by a sentence boundary.

The investigation presented above demonstrates that PA as maintained reference subject is highly correlated with the beginning of the sentence, a unit in the segmentation of discourse, which the speaker or writer decides to present as a coherent group of phrases.

The remaining 18 instances of the 144 cases of PA as maintained reference subject occur in miscellaneous sentence-internal positions. Among them 7 belong to the same pattern: they all occur in the nuclear clause after an adjunct clause with a coreferential subject encoded in ZA. Consider the following example:

(5) a. Nongren ... caǐ le yi da dui guozi, farmer pick PFV one big pile fruit

b. ☞ ranhou xialai. then come:down

c. ☞ Xialai yihou, come:down after

d. ta yi su... he one count

a. "The farmer ... picked a large pile of fruit,

b. then came down.

c. After (he) came down,

d. he counted....

(S-5)
The subject in (5)a. maintains reference through (5)a., (5)b., and (5)c. The last two clauses are presented as a separate sentence in which (5)c. serves as the adjunct to (5)d. In this sentence, the subject of (5)c. at the very beginning of the sentence is encoded in ZA, while the subject in (5)d. receives PA encoding although it is not immediately preceded by the sentence boundary. Situations like this came under examination in §4.543 in connection with the use of ZA in positions like the subject slot of (5)c. I maintain that, as compared with ZA, the choice of PA is to be attributed to the high noteworthiness of the referent.

In neither (5)c. nor (5)d. does the identification of the subject referent pose any difficulty. If continuity within sentence boundary were taken to be the only factor triggering ZA and discontinuity as marked by the sentence boundary were considered to be the only factor triggering PA, we would be in no position to explain cases like (5), which is just the opposite of the prediction. As discussed in connection with the use of ZA, sentences like (5) are by no means rare and cannot be brushed aside as accidental deviations. Instead, they constitute an important category which must be accounted for if our theory of anaphoric choice is to be descriptively adequate.

As argued in §4.543, the use of ZA in clauses like (5)c. is accounted for by the second factor underlying the choice for ZA, high negligibility. Now, we claim that the use of PA in (5)d. is to be attributed to the second factor triggering PA, high noteworthiness. Detailed discussion of the issue is given later in §4.65.
4.63 **PA for Switched Reference Subject**

When the pronominal system can differentiate the anaphor as switched reference subject from its competing referents in the discourse, naturally the subject referent can receive PA encoding, as exemplified by the following sentences:

(1) a. **San ge ren ...... jiu zou guo ne ge**
three CL person then walk past that CL
	nongchang zhuren de ne ge shu xia.
farm owner NOM that CL tree down

b. **Ta ne ge shihou zhengzai gongzuo.**
he that CL time DUR work

a. "Then the three men walked past the tree of the owner of the farm."
b. He was working at that time."  
(S-7)

(2) a. **San ge nanhai kandao...... de ge nanhai de**
three CL boy see that CL boy NOM

maozi zai di shang.
cap on ground up

b. **' Ranhou: jiu ba maozi jian qilai.**
then BA cap pick up

c. **' Ranhou: jiu cu le yi sheng koushao.**
then blow PFV one CL whistle

d. **Ta jiu hui tou yi kan.**
he then return head one see

a. "The three boys saw... that boy’s cap on the ground."
b. (They) then picked up the cap.
c. (They) then blew a whistle.
d. He then turned over his shoulders to see.  
(S-6)
In (1)b. and (2)d. above, PA ta is used to encode the subject that switches reference from the prior subject. As ta refers to singular referents only, it cannot refer to the subject referent of the preceding clause, which is the most competitive candidate for the present referent, because of the number discrepancy.

In the twenty pear stories, 25% of the switched reference subject anaphors (n = 38) receive pronominal encoding. 55% of these referents in PA (n=21) occur sentence-initially. Out of the thirty-eight PA tokens under examination, however, there are eighteen which by themselves cannot distinguish their antecedents from the competing candidates, i.e., the subject of the prior clause, within the pronominal system. Consider the following example:

(3) a. Ne wei lao xiansheng zai shu shang 
    that CL old men on tree up 
    gei le ta (Bike Boy) yi ge jihui, 
give PFV he one CL opportunity 
    b. yinwei ta zhidao.... 
because he know 
    a. "The old gentleman up in the tree gave him an opportunity, 
b. because he (Bike Boy) knew...."
   (S-14)

where ta 'he/she/it' by itself cannot differentiate the bike boy from the pear-picker. Following is another example:
(4) a. Ne jian dao mao zi de nan hai jiu ba mao zi that pick cap NOM boy then BA cap

jiao gei ta (Bike Boy),
hand give he

b. ta gei le tamen san ge bala.
he give PFV they three CL pear

a. "The boy who picked up the cap returned the cap to
him (Bike Boy),
b. he (Bike Boy) then gave them three pears."

(S-6)

where the PA in (4)b. by itself can refer to either the paddle boy or
the bike boy.

Again, such uses of PA pose an interesting question for us. Since
the PA here itself doesn’t provide enough differentiating information,
obviously the addressee must depend upon other sources of information
to insure that the addressee will make the correct inference in the
identification. Again, it is part of our goal to reveal the inference
process on the basis of the linguistic and extralinguistic knowledge
that allows the use of PA in spite of the apparent misunderstanding it
could have caused.

Here, as in the case with ZA encoding switched reference subject,
the whole gamut of linguistic and extralinguistic knowledge of the
participants of the discourse is involved in the inferential process.
For illustration, I have identified three situations in which PA is
allowed in spite of the lack of differentiating features by itself.
In the first situation, the competing referent can be deprived of its candidacy through incompatibility of the selectional features of the arguments and the predicate of the clause or through the appearance of the referent elsewhere in the clause in explicit encoding (n=9). Consider the following example:

(5) a. Ta jiu yimmian guolai,
    she then opposite come

b. ta jiu hui tou kan ne ge muhai.
    he then return head watch that CL girl

a. "At that time she came from the opposite direction,
b. and then he looked over his shoulder to watch the girl."

(5-20)

The subject of (5)a., the girl, assumes the encoding of NA in the object slot of (5)b., thus disqualifying itself for the referent of ta in the subject position of (5)b.

In the second situation, the clause with the switched reference PA subject is adjacent to the clause with the coreferential subject at the same level in the hierarchy of the rhetorical organization of the discourse, although they are separated by some other clauses in the linear order of the discourse (n=5). The situation is similar to the use of ZA as switched reference subject that pops across the prior clause back to the clause with the coreferential subject. In the case of PA, the anaphora can pop across longer distance than in the case of ZA. Consider the following example:
(6)  
   a. Ta (Bike Boy) kandao...
      he see
   b. ne ge zhuren gongzuo de taizhuanxin le,
      that CL owner work too attentively CRS
   c. dagai dou mei zhidaotala.
      probably even not know he come
   d. Ta jiu xiang yi zheng lan dou ban zou.
      he then want one whole basket all move away

   a. "He saw that...
   b. the owner was too occupied with the work,
   c. Maybe (he) didn’t even know that he came.
   d. He wanted to take away the whole basket."
      (5-16)

where (6)b. and (6)c. are embedded in (6)a., and (6)d. is adjacent to
(6)a. at the same level in the hierarchy of the discourse.

In the third situation, the clause with the PA as the switched
reference subject is so closely united with the preceding clause that
they are considered as constituting a coherent chain of actions
through involvement with a single goal (n=17), as exemplified by the
following clauses:
(7) a. Ne ge ren ba maozi na gei ta (Bike Boy) yihou, that CL person BA cap take give he after

b. ta (Bike Boy) jiu song ta ne ge san ge shuiguo. he then give he that CL three CL fruit

a. "After that person gave him (Bike Boy) the cap,
b. he (Bike Boy) gave him three pieces of fruit."

Apart from coherence achieved through involvement with a single goal of cap-pear exchange, (7)a. and (7)b. are also characterized by the adjacency of two morphologically identical PAs, one after gei 'give' in (7)a., and the other after song 'give' in (7)b. We claim that the speaker assumes the following identification strategy on the part of the hearer based on the two kinds of characteristic features provided by the linguistic context:

First, since the conventions for the verbalization of the actions involved with a single goal postulate that ZA will be triggered for the following subjects if they are coreferential with the prior one (with exceptions discussed earlier), it is not likely that the speaker would deviate from the norm without good reason. When PA is chosen for the subject of the second clause which is united with the first clause through involvement with a single goal, the subject referent must be a switched reference one, otherwise ZA would be expected.
Second, since the use of PA assumes that the antecedent is high in predictability, it must be closely at hand in the discourse. The non-subject PA in the immediately preceding clause thus stands as the most likely candidate for the anaphora.

The speaker assumes that the first step in the identification strategy as given above will lead the hearer to reject the subject referent of (7)a. as the antecedent of the subject of (7)b., and the second step will lead him to identify te after sei in (7)b. as coreferential with the subject PA in (7)b.

Note that the total number of instances in the three situations listed above exceeds eighteen. The reason for this is that the three situations may overlap in some cases. For instance, consider the following example:
(8) a. Ranhou, ne jiandao maozi de nanhaizi jiu ba then that pick cap DE boy then BA
    maoza jiao gei ta (Bike Boy),
cap hand give he
    b. ta gei le tamen san ge bala.
he give PFV them three CL pear

    a. "Then, the boy who picked up the cap handed the
cap to him (Bike Boy),
    b. he gave them three pears."

(S-14)

The boy who picked up the cap is disqualified for the antecedent of
the PA subject in (8)b. when he is included in the threesome as the
recipient of the three pears in (8)b. Moreover, the involvement of the
two clauses with the same cap-pear exchange goal makes it unlikely for
the subject of (8)b. to receive PA encoding if it is coreferential
with the subject referent in the prior clause. As a result, ta in the
postverbal slot in (8)a. is easily established as the antecedent of
the PA subject in (8)b. The cooccurrence of two or more factors for
disambiguation and identification of the anaphor as in (8) explains
why the total number of the instances in the three situations exceeds
eighteen.
4.64 PA Elsewhere in the Clause

Basically the same principles apply to the use of PA in other syntactic slots in the clause as apply in the subject position. Here, for illustration, let's consider the use of PA in the object position of the clause.

When the antecedent of the anaphor and any competing candidate can be differentiated within the pronominal system, PA may be used if the antecedent is high in continuity. Consider the following example:

(1) a. Ta (Bike Boy) yi zhan qilai de shihou, he once stand up NOM time

b. 他 kandao san ge nanhai zi zai ta mianqian a, see three CL boy at he front PAR

c. ne san ge nanhai zi man haoxinde bang ta that three CL boy quite kindly help he

ba bala jian qilai.
BA pear pick up

a. "When he (Bike Boy) stood up,
b. (he) saw three boys in front of him,
c. the three boys were kind enough to help him pick up the pears."

(S-16)

where the bike boy is encoded in PA as a possessive in (1)b., and as a pivotal object in (1)c. The only other referent in the current
discourse, the three boys, is excluded as a candidate for the reference of the PA in both positions because of the number discrepancy.

In the case where the PA by itself is not able to distinguish the anaphor from competing candidates in the context, it may still be chosen when other linguistic or extralinguistic information leads to easy identification of the antecedent. The competing referent can be rejected through other means such as explicit encoding in terms of NA. As a result, no ambiguity is involved if there is only one plausible candidate left over for the PA in the discourse. Consider the following example:

(2) a. Wu Taiji wei biao tongqing lai kan Zhang Dage,
    Wu Taiji for express sympathy come see Zhang Dage

    b. Zhang Dage meiyou jian ta.
    Zhang Dage not see he

a. "Wu Taiji came to see Zhang Dage to express his sympathy,
b. but Zhang Dage didn't see him."
    (Lao She - LH)

where the NA for Zhang Dage in (2)b. excludes it as a candidate for the reference of the PA in the same clause. In this case, the PA in (2)b. can only be inferred as referring to the other referent Wu Taiji in (2)a.
As we mentioned earlier, there is one feature that seems to be peculiar to Chinese. PA of inanimate reference is scarcely attested in Chinese. It is still less found in postverbal positions. Let's consider the following made-up clauses:

(3) a. Ni bu shu xiang zhao ben shu kankan ma?
    you not be want find CL book read PAR

    Wo ba zhe ben shu gei ni zemeyang?
    I BA this CL book give you how:about

-- b. Wo bu yao ta.
    I not want it

-- a. "Don't you want to find some book to read?
    How about this one?
    -- b. I don't want it."

Although (3)b. above is grammatical, ZA is much more preferable than PA for the object: As mentioned earlier, in the twenty pear stories, instances are rarely found in which PA encodes referents of low inherent or plot saliency. Table 4:8 shows that not a single instance of PA for inanimate referents like the book in the above sentences has been found in the object slot as in (3). This further supports the claim that the choice of PA is adversely affected when the referent is high in negligibility as defined earlier.
4.65 Explaining Choice of PA

4.651 Preliminary

By now we have shown that there are two types of PA use in Chinese discourse. One can differentiate the anaphor from competing candidates in the discourse by means of its morphological markings. With the other, the morphological form of the PA itself cannot provide enough differentiating information. As a result, the cues offered by other linguistic and extralinguistic information in the discourse are crucial for the inference process leading to correct identification. As with ZA, a complete treatment of the topic requires that the use of PA be discussed in connection with the other two major anaphoric types in Chinese discourse. When the context needs an explicit encoding type that can itself differentiate its referent from competing candidates in the discourse, NA and sometimes PA both seem eligible for the purpose. On the other hand, when the morphological form of PA itself is not differentiating enough in the sense discussed here, it seems that ZA may just as well be used for the purpose. What we are concerned with now is the question what discourse features it is that are responsible for the preference of PA over ZA, on the one hand, or NA, on the other hand.
4.652 **Discourse Features of PA**

On the basis of the examination of all the anaphors in the data, we claim that there are two major discourse features which are the most important factors triggering the selection of PA rather than ZA or NA.

PA is chosen when the referent:

1. occurs in places which are marked by minor breaks in discourse continuity, and

2. stands high in saliency in discourse.

Next, let's discuss the two features in turn.

First, let's consider minor breaks in discourse continuity. Chafe (1979) maintains that in the verbalization of the flow of thought, speakers tend to group the smallest units of memory, the foci, into what appear in language as sentences through the involvement of the units with a single referent, or with a single schema, or with a single goal. Sentence boundaries in discourse are usually accompanied by one or more of the following factors: involvement with different referents, with separate schema, or with separate goals. The presence of any of these factors marks minor breaks in the continuity of the discourse.
PA usually occurs where the continuity of discourse sustains such minor breaks as a result of one or more of the disruptive factors listed above. As attested in the data, PA tends to be chosen at the beginning of a sentence, which is separated from the preceding sentence through involvement with separate schemas and/or separate goals, as when PA encodes the sentence-initial subject referent that maintains reference from the prior clause. Also, PA tends to be chosen when the present sentence is separated from the preceding sentence through involvement with separate referents, separate schemata, and separate goals, as when PA encodes the sentence-initial subject referent that switches reference from the prior clause. As discussed in §4.62 and §4.63, 81% of the PA in subject position occur sentence-initially (147:182).

Also, PA is usually chosen when a given clause is connected to the prior one through involvement with a single schema and a single goal, but with different referents. The two clauses can be presented in one sentence. In such case, PA is often used to encode the subject referent of the second clause which switches reference from the subject referent of the prior clause, even though the pronoun by itself cannot differentiate the anaphor from the most competing candidate, the prior subject referent, for the reference of the anaphora.

Both ZA and PA can be used for the subject referent at beginning of a sentence. In most cases, the speakers are obviously left with
their individual preferences for one anaphoric form or the other. In my data, 62% of the PA in the sentence-initial position are preceded by connective adverbials, while 47% of the ZA in the position are preceded by such expressions. This shows that the insertion of those connective devices between clauses is more likely to correlate with PA than ZA for the subject of the following sentence. This is in conformity with the results of experiments reported in Chen (1984), which demonstrate that people more often have PA than ZA when the subject referent is preceded by such adverbials.

When the subject switches reference from the subject in the prior clause, PA is predominantly more frequent than ZA no matter whether the two clauses are presented in the form of one sentence, or separated by sentence boundaries. This indicates that the first major factor that triggers the choice of PA over ZA is the minor break in discourse continuity mainly caused by the switch of reference for the subject between clauses. The question of how the choice between PA and NA in such situations is made will be addressed in the next section.

The second discourse factor that underlies the choice of PA over ZA is the high noteworthiness of the referent in discourse. Other conditions being equal, the lower in noteworthiness an anaphor stands, the less likely is PA chosen for its encoding than ZA. The high vs. low noteworthiness of the referent is indicated in two major kinds of manifestations.
First, the referents which are low in inherent or plot saliency seldom receive PA encoding in discourse. The evidence for the scarcity of PA encoding for referents of low inherent saliency comes from the comparison of the percentage of the inanimate pronominal anaphors in the total instances of PA and the percentage of the inanimate nominal anaphors in the total instances of NA in the twenty pear stories: As Table 4:7 and Table 4:8 show, among all the instances of PA in the syntactic slots, only 2% encode inanimate anaphors (10:418), while among all the instances of NA, 30% encode inanimate anaphors (149:492). This indicates how rarely referents low in inherent saliency receive PA encoding in Chinese discourse.

The same is true for referents low in plot saliency, which usually indicates the importance of the role played by the referent in the plot development of the discourse. As discussed in §2.44, referents of low plot saliency are usually introduced into discourse in indeterminate encodings. As compared with referents of high plot saliency, those referents receive many fewer mentions later in the discourse. Among the 230 referents introduced in the twenty pear stories, there are 64 referents encoded in bare form, a specific type of indeterminate encoding. Later in the discourse, I have found only one anaphoric mention of an indeterminately coded referent that receives PA encoding. The other anaphoric mentions assume either ZA encoding, when the identification is easy, or NA encoding, when the identification is difficult.
Second, high vs. low noteworthiness is manifested by the position of the anaphor in the nuclear vs. adjunct clause. As detailed in the discussions of the use of ZA in §4.543, I maintain that the adjunct clause, together with all its constituent parts, is usually peripheral to the communicative goal of the discourse involved, as compared to the nucleus to which it is adjoined. Consequently, the adjunct clause is lower in noteworthiness, or higher in negligibility, than the corresponding nuclear clause. Other conditions being equal, referents are more likely to receive PA instead of ZA encoding in nuclear clauses than in adjunct clauses.

4.67 Summary

In the above discussion, we have identified two factors, location at a minor discontinuity of discourse, and high noteworthiness, as characteristic of anaphors that receive PA encoding in discourse. It is demonstrated how the evidence drawn from our data has attested to the high correlation between the choice of PA and each of the two factors. Next, let's consider the use of NA in discourse.
4.7 Use of Nominal Anaphora

4.7.1 Preliminary

NA is the third major anaphoric type that comes under detailed investigation in the present studies. As the brief discussion in §4.32 indicates, it is the anaphoric device that is more likely to be chosen as the distance between the anaphor and the antecedent expands and the number of interfering referents increases. Generally speaking, the observation is true, not only in Chinese, but also in other languages (cf. Givon 1983a).

However, there is one important question that remains to be addressed. As Tables 4:2, 4:3, and 4:4 indicate, there are quite a few instances of NA that cannot be explained in the distance-interference theory, such as the cases in which the NA is adjacent to its antecedent and there is no interfering referent between. For instance, consider the following clauses:
(1) a. Ta (Bike Boy) jiu ba zheng lou guozi dou 
    he then BA whole CL fruit all 
    ban zou le, 
    move away CRS 

b. fang zai ta jiaodache de qian zuo shang 
    put on his bike NOM front seat up 
    ban zou le. 
    move away CRS 

c. Neme, zhe ge xiao hai wang qian qi de shihou, 
    then this CL kid towards front ride NOM time 

d. fai zu lu shang peng dao ling wai yi ge mu hai zi. 
    on road up come: across other one CL girl 

a. "Then he (Bike Boy) took away the whole basket of fruit, 
b. put (it) on the front seat of his bike. 
c. Then, when the kid rode on, 
d. (he) saw another girl on the road."

(S-5)

where the subject anaphor in (1)c. maintains reference from the prior clause, and there is no other referent inserted between (1)b. and (1)c. that may compete for the subject reference in (1)c. Nevertheless, NA is chosen here instead of PA or ZA.

In this section, I will conduct an in-depth analysis of all the tokens of NA found in the data to reveal the principles underlying the choice of NA that will account for its use in situations like (1) as well as in other situations. First, I will investigate the use of NA for maintained and switched reference subjects. Then I will discuss briefly the use of NA elsewhere in the clause. Finally, I will summarize the features characteristic of the choice of NA in discourse.
4.72 NA for Maintained and Switched Reference Subject

Previous studies on anaphora in discourse have suggested that NA is usually chosen to encode anaphors that occur at major breaks in the flow of discourse, which are marked by conspicuous hesitations in speaking, and by paragraph breaks in writing (cf. Tai 1978, Hinds 1978, Chafe 1979, 1980b, 1984). Here, I use the term paragraph boundary to indicate such major breaks in discourse continuity.

As explicated in Chafe (1979), the paragraph boundary occurs at points where there are significant breaks in the coherence of space, time, topic, event, or world. Spatial coherence is broken when an action occurs in a place different from the previous one. Actions characterized by temporal coherence occur in inexorable succession, and the coherence is affected when the temporal succession is disrupted, vague, or irrelevant. Topic coherence is broken when the actor switches reference from the previous one. And event coherence is marked by the fact that all the events involved are enabling or carrying events for a climactic central event that forms the peak of an episode or else are aftermath events leading away from such a peak. Coherence in world is broken when the speaker makes shifts between two worlds, the 'real' world with the movie and the interviewer, and the imaginary 'film' world, in which there was a man picking pears, etc. For the most of the time, the speakers are in the film world.
The effect of paragraph boundaries is most conspicuous when discontinuity is registered on all the five factors mentioned above. However, in most cases, the five factors of coherence don't coincide with each other. When a change in coherence is particularly strong with one or more factors, it may manifest itself in the form of paragraph boundary in spite of the coherence with the other factors between clauses separated by the boundary.

What we are interested in here is the issue of how the choice of NA is correlated, in the place where the anaphor occurs, with the value of the five factors which may lead to paragraph boundaries. A total of 147 tokens of NA were found in the twenty pear stories encoding referents in the subject position. I have examined the discourse in which the NA occurs to establish the correlation between the anaphoric choice and the discontinuity of the five factors listed above. Table 4:13 presents the results of the investigation:
<table>
<thead>
<tr>
<th>Factor of Discontinuity</th>
<th>Number of Token of NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space/Time/Topic/Event</td>
<td>24</td>
</tr>
<tr>
<td>Space/Time/Event</td>
<td>5</td>
</tr>
<tr>
<td>Space/Topic/Event</td>
<td>13</td>
</tr>
<tr>
<td>Time/Topic/Event</td>
<td>22</td>
</tr>
<tr>
<td>Time/Topic</td>
<td>5</td>
</tr>
<tr>
<td>Time/Event</td>
<td>5</td>
</tr>
<tr>
<td>Topic/Event</td>
<td>9</td>
</tr>
<tr>
<td>Topic</td>
<td>36</td>
</tr>
<tr>
<td>Event</td>
<td>7</td>
</tr>
<tr>
<td>World</td>
<td>17</td>
</tr>
<tr>
<td>NONE</td>
<td>4</td>
</tr>
</tbody>
</table>

Total: 147

**TABLE 4:13**

Correlation Between NA and Five Factors Registering Discontinuity

As we see from the table, 24 anaphors receive NA encoding in places that are separated from the preceding discourse by virtue of discontinuity of space, time, topic, and event, and there are 36 anaphors in NA encoding where the topic coherence is broken but not the spatial, temporal, or event coherence. Following are some examples illustrating the situations.

First, let's consider the choice of NA for the anaphor where the speaker shifts from one world to the other:
(1) a. Zhe san jiao hai gaogao xing xing de yi ge ren
   this three CL kid happily one CL person
   na zhe yi ge guozi zou guoqu,
   take DUR one CL fruit walk past

b. wo xiang,
   I think

c. zhe ge nong ren keneng xin li juede,
   this CL farmer probably heart in feel

d. er, shao le yi lou guozi.
   uh less PPV one CL fruit

a. "The three kids passed happily, each with
   a piece of fruit.

b. I think,

c. the farmer maybe felt

d. uh, one basket of fruit disappeared."

(S-5)

The speaker, who was in the film world when she produced (1)a.,
shifted to the ‘real’ world with (1)b., (1)c., and (1)d., which she
uttered as an outside observer. NA is chosen for the subject anaphor
of (1)c. In the data, we have 17 tokens of NA which occurs at a point
of the world shift like this. Next, let’s look at the examples
illustrating some other situations:

232
(2) a. Jieguo ne san ge xiaohaizi, xiaonanhai, in:the:end that three CL kid boy  
jixu wang ta yuanlaizou laide  
go:on to he at:the:beginning walk come NOM  
lu shang zou,  
road up walk

b. jieguo, shi chi bala ma.  
in:the:end eat pear PAR

c. Jieguo ne ge nongfu cong shu shang xia lai,  
in:the:end that CL farmer from tree up down come

d. shi yi kan,  
one look

e. zemne shao le yi lan?  
how less PFV one CL

a. "In the end, the three kids, boys, went on along the  
road where he came from,  
b. in the end, (they) ate the pears.  
c. In the end, the farmer came down from the tree,  
d. (he) looked around,  
e. how could one basket disappear?"

(S-1)

where the four factors, space, time, topic, and event, concur to point  
to a break in (2)c., where NA is chosen to encode the subject anaphor  
the farmer.
(3) a. Ne san ge nanhai jingguo ne ge nongren de
that three CL boy pass that CL farmer NOM
shu dixia.
tree under

b. Eng, ne ge shihou ne,
uh that CL time REx

(c. nongren yijing xia lai le.
farmer already down come CRS

a. "The three boys passed below the farmer's tree.
b. By that time,
c. the farmer had already came down."

(S-20)

where the passing of the threesome and coming down of the farmer are involved with the same space. However, they are separated by discontinuity in time, topic, and event. NA is chosen to encode the farmer in (3)c.

In addition to the calculations presented in Table 4:13, I have assessed how each of the four factors, space, time, topic, and event, is correlated with the choice of NA in the 147 instances found in the pear stories. The results are presented in Table 4:14:
<table>
<thead>
<tr>
<th>Discontinuous Factor</th>
<th>Number of NA Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic</td>
<td>109 (74%)</td>
</tr>
<tr>
<td>Event</td>
<td>85 (58%)</td>
</tr>
<tr>
<td>Time</td>
<td>61 (41%)</td>
</tr>
<tr>
<td>Space</td>
<td>42 (29%)</td>
</tr>
</tbody>
</table>

**TABLE 4:14**

Correlation Between NA And Each of Four Factors Registering Discontinuity

As shown by the table, 74% of the tokens of NA co-occur with the break of topic coherence, which is the highest among the four factors. The lowest is the break of spacial coherence, which accounts for only 29% of the NA occurrences.

From the results above, we conclude that the subject referent receives NA encoding when the continuity of the clause containing the anaphor is disrupted through discontinuity of one or more factors from space, time, topic, event, or world. As Table 4:13 indicates, over half of the occurrences of NA (56%) in the subject position appear in places which are characterized by concurrence of two or more disruptive factors. As discussed in Chafe (1979), such concurrences of discontinuous factors mark major breaks in discourse continuity and usually lead to recognition of paragraph boundaries in the verbalization of the flow of thought. Thus, it can be established that

235
NA is usually chosen where the anaphor is separated from the preceding discourse by a paragraph boundary.

Moreover, Table 4:13 also indicates that NA also encodes the subject anaphor of the clause which maintains coherence with the preceding discourse through coherence of all the relevant factors except one, which is considered to be so strong as to trigger the use of NA. For instance, in the pear stories, we have found 36 instances of NA in places that are characterized by topic discontinuity only, and 17 instances of NA where the discourse continuity is disrupted through shift between the 'real' world and the film world. As Table 4:14 demonstrates, out of the factors under investigation, discontinuity of topic reference shows the strongest correlation with choice of NA for the referent in the subject position, with discontinuity of event coming second.
4.73 NA Elsewhere in Clause

As is evident in the tables from 4:2 through to 4:8, NA is much more miscellaneous in the distribution among various non-subject slots and among various semantic categories. In most cases, NA is chosen for anaphors in these positions when the other two anaphoric types are either inappropriate or impossible. Broadly, they can be grouped into two situations.

In the first situation, NA is the only anaphoric type that can indicate the identity of the anaphor in an unambiguous way. This will be necessary when neither the linguistic nor the extralinguistic context provides sufficient information for the identification process for ZA or PA to be chosen. In many cases, the use of NA encoding disqualifies the anaphor as a potential competing candidate for the referent of another attenuated anaphor that is co-present in the discourse. Consider the following example:
(1) a. Ta (Bike Boy) kandao yi ge xiao nuhai, 
    he see one CL little girl

b. suoyi ta jiu yimian guolai. 
    so she then opposite come

c. Suoyi ta jiu huitou kan na ge nuhai. 
    so he then turn:head see that CL girl

d. Ø Ranhou meiyou zhuyidao....... 
    then not:have notice

a. "He saw a little girl, 
    b. so she came in the opposite direction. 
    c. So he looked over his shoulder to watch her. 
    d. Then (he) didn't notice...."

(S-20)

If the girl in (1)c. were encoded in terms of PA or ZA, there would be 
no easy way to identify the referent in the subject and the object 
slot of the clause on the basis of the information available in the 
discourse. Following is another example:
If Yang hadn’t assumed NA encoding in (2)d., it would be very difficult to determine the referent of the ZA in (2)b. and (2)c. and the PA subject in (2)d.

In the second situation, both PA and NA are equally eligible for the anaphor so far as the identification of the referent is concerned. As we have remarked in the last section on the use of PA, however, in Chinese PA is seldom used to encode referents in low inherent or plot saliency. As a result, NA is usually considered to be the only appropriate anaphoric device for these referents when the identification of the referent is not very easy; otherwise, ZA is usually chosen. Consider the following examples:
(3) a. Ta nege ... you ge tizi,  
he that have CL ladder

b. ʈʂ da zai shu shang.  
lean on tree up

c. ʈʂ Cong tizi shang pa xialai.  
from ladder up climb down

a. "He, uh, ... had a ladder,
b. (which) leaned against the tree.
c. (He) climbed down from the ladder."

(S-6)

(4) a. Ne ta zai zai de shihou,  
then he DUR pick NOM time

b. diao xialai yi ge.  
fall down one CL

c. Ta jiu ba ne ge diao xialai de jian qilai.  
he then BA that CL fall down NOM pick up

a. "Then when he was picking,
b. one (pear) fell down.
c. Then he picked up the fallen one."

(S-2)

No ambiguity would result if ta 'it' replaced tizi 'ladder' in (3)c.,
or ne ge diao xialai de 'the fallen one' in (4)c. However, as noted
above, PA is usually chosen for referents of high saliency, and is
rarely used for the inanimate referents of low saliency as that in
(3)c., and (4)c. The evidence comes from the small percentage of PA
encoding such referents as presented in Table 4:8.
4.74 Explaining the Choice of NA

The major discourse-pragmatic factor that determines the choice of NA in Chinese is the low continuity of the referent in discourse, which is manifested in one of two situations. First, it displays itself in the nonidentifiability of the referent. Second, it displays itself in the low continuity of the discourse in which the anaphor occurs as a result of disruption of one or more discourse-integrating principles, such as involvement with the same space, time, event, or world, with the preceding discourse. Now let's consider the first situation.

The primary function of anaphora is relating a given referent to an entity that has already been present in the discourse so that coreference can be established between the two in the addressee's mind. As reiterated in the discussion of the identification of ZA or NA in the above sections, in addition to the forms assumed by the anaphora per se, there exists in discourse other linguistic and extralinguistic information which may lead to unambiguous identification of the anaphor when the anaphoric form doesn't provide such identifying information by itself. However, such linguistic and extralinguistic information may not always be sufficient for the identification, if the anaphora assumes an attenuated form such as ZA or PA. Consider the following sentence:

241
(1) a. Ne xiaohaizi jiu kan le ne ge nu de yi yan. that kid then see PFV that CL female NOM one eye

b. Ranhou the ge shihou che jiu pengdao then this CL time bike then bump

   pangbian yi kuai shitou. side one CL stone

c. The ge chezi jiu dao le. this CL bike then fall CRS

d. Ne ge shuiguo jiu san le man di. that CL fruit then scatter PFV over ground

a. "The kid glanced at the girl.
b. Then at this time the bike bumped against
   a stone by the roadside.
c. The bike fell.
d. The fruit scattered over the ground." (S-7)

If the underlined NP anaphors were replaced by ZA or PA in the above clauses, the speaker would fail to convey what is going on in the episode. The nonidentifiability of these anaphors as those in (1) can be attributed to many factors. In most cases, it is a result of the switch of reference from the prior subject and lack of sufficient linguistic and extralinguistic information to signal the switch, as illustrated by (1). In such situation, NA is the only choice for the anaphor.

The referent may also be separated from the preceding discourse through the discontinuity of the space, time, event, or world in which the the clause containing the anaphor is involved, as discussed in §4.72. In such a situation, NA may be chosen when the anaphor occurs
in places marked by one or more of the these disruptive factors, even if the identification of the referent poses little difficulty. These places are often identified as paragraph boundaries in discourse.

Discontinuity of space, time, event, or world coincide with discontinuity of referent. Depending upon the number and extent of the disruptive factors present in the discourse, continuity of referent constitutes a continuum, which provides the addressers with variations between the anaphoric choices. Thus, they can have their personal preferences for one anaphoric form or another on the basis of their own judgement about the continuity of the discourse. The more discontinuous the referent is in their judgement, the more likely it is to receive NA encoding.

4.8 Summary

We began this chapter with a brief discussion of the major anaphoric types in Chinese. Later on, we presented a general picture of their distribution in the twenty pear stories from three perspectives: discourse-pragmatic, syntactic, and semantic. After that, we conducted an in-depth analysis of each of the three anaphoric types to reveal the basic principles underlying the choice of each of them in discourse. The following chapter is a summary of what we have found concerning the referent introduction and referent tracking in Chinese narrative discourse.
Notes to Chapter 4

1 It must be pointed out that Japanese does have pronominal anaphora although none was found in Clancy (1980)'s spoken corpus.

2 The difference between ZA and PA is not very significant here. Nevertheless, the fact that ZA in Chinese is not longer in persistence than PA is not in conformity with the findings reported in Givon (1983a).

3 We may also consider parallelism to be the basic principle that contributes to the identification of a maintained reference subject as well as a switched reference subject discussed here.
CHAPTER 5

CONCLUSION

In the previous chapters, we have investigated how referents are introduced into Chinese narrative discourse and how they are tracked throughout the discourse in terms of various anaphoric devices.

The study shows that on the first mention of a referent in discourse it may receive one of three lexical encodings in two broad categories: determinate (indefinite or definite), and indeterminate. The choice among the three reflects assumptions on the part of the speaker about the identifiability of the referent for the listener. Those encoded in indefinite terms are assumed to be nonidentifiable. Those in definite terms are presented as if the listener can identify the referents in the context although it is their first explicit mention in the discourse. There is no such categorial correlation between indeterminate encodings and assumptions about identifiability. For referents that receive indeterminate encodings, the syntactic slots they occupy can usually be taken as a cue to the assumption that
the speaker holds about the state of identifiability of the referent for the listener.

Data analysis leads to the conclusion that the choice of the encodings for referents on initial mention depends heavily on the saliency of the referents in discourse. Indeterminate encodings in Chinese tend to be chosen for initial-mention referents that are low in saliency, while determinate encoding (definite and indefinite) is chosen for referents high in saliency. Definite encoding tends to encode the referents which are judged to be high in plot saliency, but low in inherent saliency, while indefinite encoding tends to encode those that are high in either inherent or plot saliency.

The referents are subject to one of the three major types of anaphoric encodings, ZA, PA, and NA, after they have been introduced into the discourse. ZA is considered as a linguistic anaphoric form in the sense that it constitutes a hole in the clause where an anaphor is understood without being explicitly expressed. Extensive use of ZA is a characteristic feature of Chinese that distinguishes it from Indo-European languages like English. PA, because of the pronominal form itself in Chinese, carries much less differentiating information than PA in languages like English. Thus, in most situations in which such attenuated anaphoric forms are used, various other types of linguistic and extra-linguistic information present in the discourse play a crucial role in the identification of the anaphora through inference.
The choice of a specific type of anaphoric encoding depends upon the identifiability and the notworthiness of the referent, and the continuity of the discourse in which the anaphor occurs with the preceding discourse that contains its antecedent. In general, since ZA and PA in Chinese carry little or no differentiating capacity, the presence of sufficient linguistic and extralinguistic information which can lead to unambiguous identification of the anaphor constitutes a prerequisite to the use of ZA or PA in narrative discourse. Given this precondition, ZA is used when the referent occurs in discourse that is high in continuity as assessed in terms of the several discourse-integrating principles we have mentioned or when it is high in negligibility as indicated by its low plot saliency or the adjunct status of the clause that contains it; PA is used when the referent occurs in discourse marked by minor breaks and the referent is high in notworthiness. On the other hand, NA is used when the referent is difficult to identify due to the distance and/or interference between the anaphor and its antecedent, or when the discourse containing the referent is marked by major breaks in its continuity to the preceding discourse.

On the basis of my preliminary investigation, I have reasonably strong evidence to suggest that the principles underlying the anaphoric choice in Chinese apply to other languages that are characterized by the inference system as well. As a case study, the findings reported in the present investigation lead to a deeper understanding of the the whole inference system in the typology of
referent-tracking mechanisms offered by Foley and Van Valin (1984). Further inquiries along the similar line into the other languages will prove to be very helpful in working out a complete picture of the inference system, at the same time contributing to a better knowledge of the whole gamut of the mechanisms under investigation.

Moreover, it is generally recognized that the heavy use of ZA constitutes a distinctive feature of the grammar of the languages characterized by the inference system. The fact that the use of ZA as well as other anaphoric devices depends heavily upon discourse-pragmatic information demonstrates beyond any doubt how this important demain of grammar must be understood and explained from a discourse-pragmatic perspective.
BIBLIOGRAPHY


In: Givón (1983a).

Vol. 1, Amsterdam: John Benjamins.


