Getting to Know Referring Expressions: Anaphor and Accessibility in Mandarin Chinese[1]

Shuanfan Huang

National Taiwan University

*In order to do pronoun resolution, one
has to be able to do everything else.*

Charniak (1972)

1. Introduction

Referring expressions are used to refer to entities in the world directly or independently. They can also be used to refer in a way that depends on some linguistic antecedents for referent interpretation. These latter referring expressions are called anaphors and the entities they refer to are antecedents. How does a hearer know exactly which antecedent a particular anaphoric expression refers to? This is the problem of anaphor resolution. Anaphor resolution has become something of a growth industry for cognitive scientists interested in an abstract characterization of the syntactic phenomenon involved or in modeling thought processes and developing natural language processing systems. Anaphora has proved to be well-suited as a testing ground for such enterprises since attempts to model anaphor resolution have shown that it involves a surprisingly large number of variables: lexical, morphological, syntactic, semantic, pragmatic and even suprasegmental.[2] Charniak (1972) cited in Hobbs (1978) was the first researcher to demonstrate that in order to do pronoun resolution, one had to be able to do everything else. In linguistics, while many linguists continue to toy with the idea of providing a formal account of constraints on intrasentential anaphora, Keenan (1976), Bolinger (1977, 1979), Li and Thompson (1979) were among the first scholars to espouse the radical view that there are no structurally statable restrictions on anaphora and that attempts at improving formal accounts are often mere ad hoc patchworks on the basic syntactic analysis which does not work.

Against this background, the purpose of this paper will be to provide empirical evidence for an accessibility scale for Chinese anaphoric expressions in two modes of discourse and to justify the existence of various types of anaphors in the pragmatics of language. Following Ariel (1990), I suggest that natural languages provide speakers with means to encode the

27

accessibility of the antecedent of an anaphor to the hearer. To anticipate a major finding of this study, the accessibility scale for referring expressions in Chinese can be expressed in (1):

(1) Accessibility Scale for Chinese Anaphoric Expressions:

zero anaphor > pronoun > determinate NP > proper noun > bare NP >

possessive NP > complex NP

This means that zero anaphors mark the highest degree of accessibility and complex NPs the lowest. When the speaker uses ZA, he assumes its antecedent is the most accessible to the hearer, while use of the complex NP means its antecedent is the least accessible to the hearer.

Marker accessibility is closely tied to context types and structure types. Three context types can be distinguished (Clark and Marshall 1981; Ariel 1990): general context, physical context and linguistic context. Lexical NPs (proper nouns, bare NPs etc.) usually refer to general knowledge, demonstrative expressions to the physical context and pronouns and ZA to the linguistic context for referent interpretation. In an initial referential act, the speaker commonly uses a full lexical NP (proper noun, bare NP, or NP containing a relative clause) and the speaker refers the hearer to the most general knowledge for the interpretation of new entities associated with the lexical NP. Hence full NPs can be said to be the most 'informative' of referring expressions. However, information stored in general knowledge takes longer to activate and interpreting a referent on the basis of that type of information takes correspondingly longer and proceeds at a greater cost to working memory. By contrast, recent linguistic material is the most accessible information, as evidenced by the use of the semantically emptier or empty high accessibility markers such as pronoun or ZA on a second mention.

Structure types refer to whether the antecedent/ anaphor pair occurs in the same clause, the same sentence, the same topic chain or the same paragraph or not. High accessibility markers tend to attract their antecedents to a position within a smaller discourse unit, whereas low accessibility markers tend to find their antecedents in a more distant text position.

There is thus a three-way correlation between anaphoric form, context type and structure type. This means that factors that govern the choice of anaphoric forms involve both information having to do with the type of context in which an antecedent occurs and the distance separating an anaphor and its antecedent. Operationalized in concrete terms, choice of anaphoric expressions is determined by the following four factors (Ariel 1990):

(a) Distance: the distance between antecedent and its second-mention anaphor;

(b) Competition: the number of competitors for the role of antecedent;

(c) Saliency: topicality of first retrievals and of subsequent mentions, i.e.

whether new entities are mentioned as topics in a main

clause or not;

(d) Unity: whether the antecedent/ anaphor pair is in the same sentence,

topic chain or paragraph.

28

The relationship between the four factors and degree of accessibility can be stated as follows:

1. the more distant from its anaphor an antecedent is, the less accessible it is;
2. the more competitive an antecedent is, the less accessible it is;
3. the more salient an antecedent is, the more accessible it is;
4. the higher the unity between an antecedent and its anaphor, the more accessible it is.

Accessibility, equivalent in substance to Grosz' (1977) notion of focussing, is also called activation by Collins and Quillian (1972). In processing terms, the hearer attempts to pick as the antecedent of an anaphoric form from among a number of competitors by assessing the accessibility (or activation) of antecedents in his discourse model based on the four factors given above. When a word is recognized, the associated concept in long term memory is activated and remains in this state for some time, after which it begins to dissipate, unless reactivated at appropriate intervals. Anderson (1976) suggests that activation levels decrease in proportion to their distance from their initially activated concept. Thus the shorter the distance between an antecedent and its anaphor, the higher state of activation it is in; consequently, co-reference can be achieved by using a more attenuated anaphoric form, i.e. a higher accessibility marker. On the other hand, reactivation of an entity must be done via a full lexical NP.

2. Database

Five oral narratives about the movie 'Ghost' and five written texts formed the database of this study. There were a total of 1295 anaphoric expressions, 454 of which came from the written texts and 841 of which were from the spoken texts. These anaphoric expressions fall into the following seven types, in roughly ascending order of informativity:

a) zero anaphor

b) pronoun (including reflexives)

c) determinate NP: i) demonstrative: zheige, neige

   ii) dem plus bare NP: zheige xuesheng

   iii) dem plus proper noun: zheige Mali

d) bare NP: xuesheng, lingmei;

e) possessive NP: tade xuesheng

f) proper noun

g) complex NP: i) relative clause plus bare NP: tade nupenyou suoyang de mao

   ii) possessive NP plus proper noun: tade xuesheng Yang Jian

Both the written texts and the spoken texts are segmented into three levels of discourse units: clauses, sentences and paragraphs for the former and clauses, topic chains and paragraphs for the latter. A clause in a focus of consciousness and is roughly equivalent to a

case frame, namely, a predicate with its associated arguments. In spoken discourse, a clause is usually realized as an intonation unit (Chafe 1987: 22). A sentence is a discourse unit of local cohesion established between adjacent clauses, and is marked off from one another by periods or question marks.

A topic chain, a stretch of discourse consisting of two or more clauses sharing the same topic, corresponds roughly to the extended sentence in the sense of Chafe (1980), which though intonationally and syntactically belongs to more than one sentence, is expressed as a coherent mental image. Chafe (1980) also speaks of the notion of a center of interest, which represents a mental image and contains a set of events leading to a goal. A center of interest, then is in essence equivalent to a topic chain in the sense intended here.

A paragraph as a discourse unit can be justified by paragraph closure and thematic unity. Paragraph closure refers to the features marking the boundaries of a paragraph. Paragraph boundaries often occur at points of major shifts in scene, time, space, world, character configuration, or event structure, which require the speaker to reorient himself and thus often result in his processing difficulty, signalled by hesitations, false starts, or long pauses (Chafe 1980, 1987). In the spoken texts examined for this study, paragraph introducers like the following were found: haole 'well, OK', na 'and then', you yitian 'one day', jingtou zhuandao yi yinhang 'the scene switches to a bank.' They were usually preceded or followed by long pauses.

The written texts selected for analysis consist of a total of 39 paragraphs, 122 sentences and 341 clauses; the spoken texts comprise a total of 37 paragraphs, 135 topic chains and 467 clauses. Mean lengths of each discourse unit in terms of words, clauses and sentences are shown in Table 1.[3]

Writting	Speech
clause = 7.4 words	clause = 6.1 words
sentence = 2.8 clauses	topic chain = 3.46 clauses
paragraph = 3.13 sentences	paragraphs = 3.64 topic chains

Table 1. Mean Length of Each Discourse Unit

Table 2 gives a breakdown of the anaphoric forms found in the written and spoken texts.

Anaphoric form	writing		speech	
	N	%	N	%
ZA	218	48	185	22
Pronoun	89	19.6	292	34.7
det NP	13	2.8	155	18.4
bare NP	55	12.1	39	4.6
poss NP	7	1.5	58	6.9
proper NP	64	14	100	11.9
complex NP	8	1.7	12	1.4
total	454	100	841	100

df=6, chi-square=182.67, p<.001

Table 2 Breakdown of Anaphoric Forms in Written and Spoken Texts

Patterns of distribution of anaphoric forms in the spoken and written texts are significantly different (df=6, chi-square=182.67, $p < .001$). Much of this difference can be attributed to the fact that the written texts use a higher proportion of the more accessible anaphoric expressions - - ZA and pronouns together account for 67.6% of all the anaphoric expressions used in writing as against 56.7% in speech, and that determinate NPs are much more popular with the spoken texts. These differences can be seen as a consequence of the distinct ways speech and writing are designed. This is a topic of considerable theoretical interest, but will not be taken up in this paper.[4]

2.1. Accessibility and Referential Distance

Although popularity of the various types of anaphoric expressions varies significantly, they can be shown to observe basically the same pattern of accessibility in the two modes of discourse based on their distribution with respect to referential distance. Table 3 and Table 4 present the relevant statistics.

Anaphoric form	Text Position						
	same clause	previous clause	same sentence	previous sentence	same paragraph	across paragraph	total
ZA	17 (7.8) (56.7)	166 (76.1) (82.6)	10 (4.6) (43.4)	18 (8.3) (17.6)	3 (1.4) (12.5)	4 (1.8) (5.4)	218 (48)
Pronoun	11 (12.4) (36.7)	21 (23.6) (10.4)	7 (7.9) (30.4)	43 (48.3) (42.1)	5 (5.6) (20.8)	2 (2.2) (2.7)	89 (19.6)
demonstrative	0	1 (100) (0.5)	0	0	0	0	1 (0.2)
dem + bare NP	0	2 (16.6) (1)	0	5 (41.6) (5)	3 (25) (12.5)	2 (16.6) (2.7)	12 (2.6)
dem + PN	0	0	0	0	0	0	0
bare NP	1 (1.8) (3.3)	7 (12.7) (3.4)	3 (5.5) (13)	14 (25.5) (13.7)	6 (10.9) (25)	24 (43.6) (32.4)	55 (12.1)
Poss NP	0	0	1 (14.3) (4.3)	1 (14.3) (1)	3 (42.9) (12.5)	2 (28.6) (2.7)	7 (1.5)
Proper noun	1 (1.6) (3.3)	4 (6.3) (2)	2 (3.1) (8.6)	19 (29.7) (18.6)	2 (3.1) (8.3)	36 (56.3) (48.6)	64 (1.4)
rel + NP	0	0	0	1 (16.7) (1)	2 (33.3) (8.3)	3 (50) (1.4)	6 (1.2)
Poss NP + PN	0	0	0	1 (50) (1)	0	1 (50) (1.3)	2 (0.4)
total	30 (6.6)	201 (44.3)	23 (5)	102 (22.5)	24 (5.3)	74 (16.3)	454

Table 3 Breakdown of Anaphoric Forms in Written Texts as to Referential Distance

Anaphoric form	Text Position						
	same clause	previous clause	same topic chain	previous topic chain	same paragraph	across paragraph	total
ZA	2 (1.1) (5)	176 (95.1) (41.9)	4 (2.1) (8)	3 (1.6) (1.6)	0	0	185 (22)
Pronoun	34 (11.6) (85)	162 (55,5) (38.6)	22 (7.5) (44)	59 (20.2) (32.4)	3 (1) (5)	12 (4.1) (13.6)	292 (34.7)
demonstrative	0	3 (75) (0.7)	0	1 (25) (0.5)	0	0	4 (0.4)
dem + bare NP	0	49 (37.1) (11.6)	9 (6.8) (18)	47 (35.6) (26.3)	14 (10.6) (23)	13 (9.8) (14.7)	132 (15.7)
dem + PN	0	0	0	5 (26.3) (2.7)	5 (26.3) (8.2)	9 (47.4) (10.2)	39 (4.6)
bare NP	0	9 (23.1) (2.1)	3 (7.7) (6)	12 (30.8) (6.6)	6 (15.4) (10)	9 (23.1) (10.2)	39 (4.6)
Poss NP	1 (1.7) (2.5)	7 (12.1) (1.7)	1 (1.7) (2)	23 (39.7) (12.6)	14 (24.1) (23)	12 (20.7) (13.6)	58 (6.9)
Proper noun	3 (3) (7.5)	14 (14) (3.3)	11 (11) (22)	32 (32) (17.6)	17 (17) (27.8)	23 (23) (26.1)	100 (11.9)
rel + NP	0	0	0	0	2 (28.6) (3.3)	5 (71.4) (5.7)	7 (0.8)
Poss NP + PN	0	0	0	0	0	5 (100) (5.7)	5 (0.6)
total	40 (4.7)	420 (50)	50 (5.9)	182 (21.6)	61 (7.2)	88 (10.5)	841

Table 4 Breakdown of Anaphoric Forms in Spoken Texts as to Referential Distance

Table 3 shows that zero anaphors have the most recent antecedents and complex NPs the most remote ones. Specifically, 96.8% of the zero anaphors find their antecedents not further than the immediately previous sentence; the corresponding ratio for pronominal anaphors is 92.2%; for determinate NPs, 61.5%; for bare NPs, 45.5%; for possessive NPs, 28.6%; for proper nouns, 40.6% and for complex NPs, 25%. We thus see that referential distance has a considerable effect on the choice of anaphoric forms. This is only natural, since the more remote an antecedent becomes, the more difficult it would be for it to remain in the same state of activatedness and to participate in cognitive processing, making referent identification that much harder for the hearer, unless the speaker provides more informative cues commensurate with the distance separating the antecedent/ anaphor pair.

Based on Table 3 we take [2.1] to be the accessibility scale for Chinese anaphoric forms in written texts based on distance.

[2.1] Accessibility Scale of Anaphoric Expressions in Writing Based on Distance:

ϕ > pronoun > determinate NP > bare NP > possessive NP > proper noun > complex NP

Table 4 shows the same basic pattern of accessibility in the spoken texts as in the written texts: zero anaphors have the most recent antecedents and complex NPs the most remote ones. Specifically, 100% of the zero anaphors find their antecedents occurring not further than the previous topic chain; for pronouns, the ratio is 94.9%; for determinate NPs (excluding dem plus PN), 80.1%; for bare NPs, 61.5%; for possessive NPs, 55.1%; for proper nouns, 60%, and for dem plus PN, 26.3%. At the extreme end of the scale, all complex NPs favor the most distant position (across the paragraph) for their antecedents. We thus arrive at [2.2] as the accessibility scale for anaphoric expressions in spoken texts based on distance:

[2.2] Accessibility Scale of Anaphoric Expressions in Speech Based on Distance:

ϕ > pronoun > determinate NP > bare NP > proper noun > possessive NP > dem plus PN > complex NP

Proper nouns modified by demonstratives (i.e. dem plus PN), found only in the spoken texts, behave like a lower accessibility marker and, in fact, are significantly less accessible than proper nouns (df=1, chi-square=7.36, p <.01). However, a comparison between bare NPs and dem plus bare NPs (i.e. bare NPs premodified by demonstratives) shows the opposite effect: the former are significantly less accessible than the latter. Why do demonstratives seem to have divergent influences on the behavior of these two types of anaphoric expressions? This apparent paradox will be resolved in Section 3.

Although the accessibility scales in [2.1] and [2.2] are nearly indistinguishable, the ways written and spoken texts structure their accessibility are significantly different. Table 5 presents the same data as those in Table 3, but collapsing references to same clause and previous clause positions with references to same sentence position (since whether an antecedent/ anaphor pair is in the same clause, or one clause apart from each other, they are all

in fact in the <u>same sentence</u>). Similarly, Table 6 presents the same data as those in Table 4, but collapsing references to <u>same clause</u> and <u>previous clause</u> with references to <u>same topic chain</u>.

Anaphoric form	Text Position				
	same sentence	previous sentence	same paragraph	previous paragraph	total
ZA	193 (88.5) (76.3)	18 (8.3) (17.6)	3 (1.4) (12.5)	4 (1.8) (5.4)	218 (48)
Pronoun	39 (43.8) (15.4)	43 (48.3) (42.1)	5 (5.6) (20.5)	2 (2.2) (2.7)	89 (19.6)
dem + bare NP	2 (16.6)	5 (41.6) (5)	3 (25) (12.5)	2 (16.6) (2.7)	12 (2.4)
bare NP	11 (20) (4.3)	14 (25.5) (13.7	6 (10.9) 25	24 (43.6) (32.4)	55 (12.1)
Poss NP	1 (14.3) (0.8)	1 (14.3) (1)	3 (42.9) (12.5)	2 (28.6) (2.7)	7 (1.5)
Proper noun	7 (10.9) (2.7)	19 (29.7) (18.6)	2 (3.1) (8.3)	36 (56.3) (48.6)	64 (14)
rel + NP	0	1 (16.7) (1)	2 (33.3) (8.3)	3 (50) (4)	6 (1.2)
Poss NP + PN	0	1 (50) (1)	0	1 (50) (1.3)	2 (0.4)
total	253 (55.8)	102 (22.5)	24 (5.3)	74 (16.3)	454

Table 5. Breakdown of Anaphoric Expressions in Written Texts in Conflated Text Positions

Anaphoric form	Text Position				
	same TC	previous TC	same P	previous P	total
ZA	182 (98.4) (35.7)	3 (1.6) (1.6)	0	0	185 (22)
Pronoun	218 (74.6) (42.7)	59 (20.2) (32.4)	3 (1) (5)	12 (4) (13.6)	292 (34.7)
demonstrative	3 (0.5)	1 (0.5)	0	0	4 (0.4)
dem + bare NP	58 (44) (11.3)	47 (35.6) (25.8)	14 (10.6) (22.9)	13 (9.8) (14.7)	132 (15.7)
dem + PN	0	5 (26.3) (2.7)	5 (26.3) (8)	9 (47.4) (10.2)	19 (2.2)
bare NP	12 (30.7) (2.3)	12 (30.7) (6.6)	6 (15.4) (9.8)	9 (23) (10.2)	39 (4.6)
Poss NP	9 (15.5) (1.7)	23 (39.6) (12.6)	14 (24.1) (23)	12 (20.7) (13.6)	58 (6.9)
Proper noun	28 (28) (5.5)	32 (32) (17.6)	17 (17) (27.8)	23 (23) (26.1)	100
rel + NP	0	0	2 (28.6) (3.2)	5 (71.4) (5.7)	7 (0.8)
Poss NP + PN	0	0	0	5 (100) (5.7)	5 (0.6)
total	510 (60.6)	182 (21.6)	61 (7.2)	88 (10.4)	841

Table 6 Breakdown of Anaphoric Expressions in Spoken Texts in Conflated Text Positions

TC = topic chain: P = paragraph

Based on Table 5 and Table 6 we construct the contingency table in Table 7. The chi-square test shows that the distributions of the anaphoric expressions in writing and speech as to text positions are significantly different (x^2=11.05, df=3, p <.05).

(a) observed frequencies

	same S/TC	previous S/TC	same P	previous P	total
writing	253	102	24	74	453
speech	510	182	61	88	841
total	763	284	85	162	1294

(b) expected frequencies

	same S/TC	previous S/TC	same P	previous P	total
writing	267.1	99.4	29.7	56.7	453
speech	495.9	184.6	55.3	105.3	841
total	763	284	85	162	1294

Table 7 Contingency Table of Anaphoric Expressions as to Text Positions

It is possible to deduce further significant differences between the two modes of discourse from Table 3 and Table 4. Examining Table 3, we observe that ZA favors a position where the antecedent occurs in the previous clause; pronouns favor a position where the antecedent occurs in the previous sentence; proper nouns and complex NPs favor the most distant position, across the paragraph; bare NPs and possessive NPs favor the same paragraph position. Finally, determinate NPs favor the previous sentence position, the difference between them and pronouns being mainly reflected in their secondary environments: pronouns being preferred in the previous clause position and determinate NPs in the same paragraph position. Secondary environments are similarly important for any two (or three) anaphoric expressions which favor the same primary text position. For example, both bare NPs and poss NPs favor the same paragraph position, but are distinguished from each other in secondary preferences: bare NPs favor the previous sentence position and poss NPs the position more remote than the previous sentence.

In Table 4, representing the spoken texts, ZA favors a position where the antecedent occurs in the previous clause, as in the written texts, but its overall popularity slipped to just 22% from a high of 48% for written texts. More significantly, pronouns now favor a position where the antecedent occurs in the previous clause; bare NPs, proper nouns and possessive NPs favor the more distant position of previous topic chain, and the most remote position, namely, across the paragraph, now is favored by complex NPs. Determinate NPs do not have a clear-cut preference, with 33.5% showing preference for previous clause position, and

35

34.2% for previous topic chain position. But since the same topic chain position, broadly construed to take in previous clause positions, accounts for 39.3% of the occurrences, I thus take the same topic chain as the favored position for determinate NPs.

Table 8 summarizes the preferred positions for antecedents for the various anaphoric expressions. It is clear that, for each anaphoric expression, its antecedent in the spoken texts consistently favor the more recent positions than those in the written texts, a consequence of the fact that the spoken language fades rapidly and hence requires successively more informative expressions to aid the hearer in referent identification, resulting in the asymmetric distribution of anaphoric expressions in the two modes of discourse.

Table 8 is a convincing demonstration of a critical difference between speech and writing as manifested in the linguistic behavior of anaphoric choice. It also provides strong evidence for a three-way distinction between high, mid and low accessibility markers based on distributional facts of referential distance -- ZA and pronouns are high accessibility markers, determinate NPs mid accessibility markers and the rest low accessibility markers.

Anaphoric form	Preferred position for antecedent	
	writing	speech
ZA	previous clause	previous clause
Pronoun	previous sentence	previous clause
determ NP	previous sentence	same topic chain
bare NP	same paragraph	previous topic chain
Poss NP	same paragraph	previous topic chain
Proper noun	across paragraph	previous topic chain
complex NP	across paragraph	across paragraph

Table 8 Preferred Positions for Antecedents of Each Anaphoric Expression in Speech and Writing

2.2. Accessibility and Competition

However, referential distance is not the only factor that distinguishes anaphoric expressions from one another. Nor is it the only factor that affects accessibility. Closely related to distance as an indication of marker accessibility is competition, an idea made popular by Givon (1983). This refers to the idea that the number of potential competitors intervening between an anaphor and its antecedent varies in inverse proportion to the hearer's ability to retrieve the intended antecedent. As competition increases over antecedenthood, the anaphoric expression must also

become more 'competitive'. But a competitive anaphoric expression necessarily contains more lexical information, other things being equal, for referent resolution to be possible. It follows that the more competitive anaphors are also the lower accessibility markers.

Table 9 and Table 10 present the distributions of anaphoric expressions with respect to competition.

Anaphoric form	Competitor					
	0	1	2	3	4+	total
ZA	180 (82.6) (67.2)	31 (14.2) (37.8)	6 (2.7) (17.6)	0	1 (0.4) (0.5)	218
Pronoun	46 (51.7) (17.2)	26 (29.2) (31.7)	13 (14.6) (38.2)	2 (2.2) (15.3)	2 (2.2) (3.4)	89
demonstrative	0	1	0	0	0	1
dem + bare NP	4 (33.3)	3 (25) (3.6)	2 (16.7)	2 (16.7) (15.3)	1 (8.3)	12
bare NP	14 (25.5) (5.2)	9 (16.4) (10.9)	6 (10.9) (17.6)	4 (7.3) (30.7)	22 (40) (37.9)	55
Poss NP	1 (14.3)	1 (14.3)	0	1 (14.3)	4 (57.1) (6.8)	7
Proper noun	22 (34.4) (8.2)	10 (15.6) (12.2)	6 (9.4) (17.6)	3 (4.7) (23)	23 (35.9) (39.6)	64
rel + NP	0	1	1	1	3 (50) (5.1)	6
Poss NP + PN	0	0	0	0	2 (100) (3.4)	2
total	267 (58.8)	82 (18)	34 (7.5)	13 (2.8)	58 (12.7)	454

Table 9 Breakdown of Anaphoric Expressions in Written Texts as to Competition

Anaphoric form	Competitor					
	0	1	2	3	4+	total
ZA	121 (65.4) (35.2)	57 (30.8) (25.1)	6 (3.2) (7)	0	1 (0.5)	185
Pronoun	132 (45.2) (38.5)	98 (33.6) (43.1)	36 (12.3) (42.3)	13 (4.5) (37.1)	13 (4.5) (8.6)	292
demonstrative	3 (75)	1 (25)	0	0	0	4
dem + bare NP	46 (34.8) (13.4)	32 (24.2) (14)	11 (8.3) (8.3)	10 (7.5) (28.5)	33 (25) (21.8)	132
dem + PN	1 (5.2)	1 (5.2)	3 (15.7)	1 (5.2)	13 (68.4) (8.6)	19
bare NP	12 (30.7) (3.5)	9 (23)	3 (7.7)	2 (5.1)	13 (33.3) (8.6)	39
Poss NP	13 (22.4) (3.8)	8 (13.8)	10 (17.2) (11.7)	4 (6.9) (11.4)	23 (39.6) (15.2)	58
Proper noun	15 (15) (4.3)	21 (21) (9.2)	16 (16) (18.8)	4 (4) (11.4)	44 (44) (26.4)	100
rel + NP	0	0	0	1	6	7
Poss NP + PN	0	0	0	0	5	5
total	343 (40.7)	227 (27)	85 (10.1)	35 (4.2)	151 (17.9)	841

Table 10 Breakdown of Anaphoric Expressions in Spoken Texts as to Competition

Table 9 shows that 96.8% of the zero anaphors appear with no or only one competitor intervening between themselves and their antecedents. The corresponding ratio for pronouns is 80.9%; for determinate NPs, 61.5%; for proper nouns, 50%; for bare NPs, 41.9%; for possessive NPs, 28.6%, and for complex NPs, 12.5%. The competition scale for the written texts is then as follows.

[2.3] Accessibility Scale of Anaphoric Expressions in Written Texts Based on Competition:

φ > pronoun > determinate NP > proper noun > bare NP > poss NP >

complex NP

Statistically, there is no significant difference in patterns of distribution between any two adjacent anaphoric expressions in [2.3] except between ZA and pronouns, though there is between any two non-adjacent anaphors. We will therefore continue to make a three-way distinction and take determinate NP as the midpoint of the scale, yielding the following accessibility scale:

high accessibility markers: ZA, pronoun

mid accessibility markers: determinate NP

low accessibility markers: proper noun, bare NP, poss NP, complex NP

The accessibility scale of anaphoric expressions in the spoken texts based on competition is as shown in [2.4], as can be derived from Table 10.

[2.4] Accessibility Scale in Spoken Texts Based on Competition:

φ > pronoun > determinate NP (excluding dem plus PN) > bare NP >

poss NP > proper noun > dem plus PN > complex NP

Proper nouns modified by demonstratives (i.e. dem plus PN) behave like a lower accessibility marker. Indeed they are significantly less accessible than proper nouns (df=1, chi-square=4.59, p <.05), exactly as in [2.2]. Moreover, bare NPs modified by demonstratives (i.e. dem plus bare NP), shown in [2.4] as determinate NP, again show greater accessibility than bare NPs, as in [2.2], though their difference is not statistically significant.

A comparison between Table 9 and Table 10 reveals a notable difference that is rooted in the difference between speech and writing we have observed in connection with Table 8: anaphoric expressions in spoken texts are significantly less competitive than those in the written texts (df=1, chi-square=11.98, p <.001). The contingency table in Table 11 presents the pertinent statistics:

(a) observed frequencies			
	low competition	high competition	total
writing	349	105	453
speech	570	271	841

(b) expected frequencies			
	low competition	high competition	total
writing	322	132	454
speech	597	244	841

Table 11 Contingency Table of Competition in Spoken and Written Texts

Competition and distance are two sides of the same coin. High competition is much more likely to correlate with longer distance and lower accessibility; low competition correlates with shorter distance and higher accessibility. If there is only one activated entity (i.e. if there is no competitor) in the hearer's discourse model, then the system can link it to an attenuated anaphor (i.e. a higher accessibility marker). But if there are two or more activated entities, the system will obviously need more informative anaphors (i.e. lower accessibility markers) for referent resolution. Since we have shown in Table 8 that antecedents for all anaphoric expressions of the spoken texts consistently favor the more recent positions than those of the written texts, it follows that anaphoric expressions in the spoken texts should be less competitive than those in the written texts, exactly the results derivable from the contingency table Table 11.

2.3. Accessibility and Saliency

A third factor affecting accessibility is saliency. This refers to the relationship between anaphoric expressions and the saliency (i.e. topicality) of their antecedents. A topical antecedent is in a higher state of activatedness and hence more highly accessible than a non-topical antecedent. Table 12 presents the relevant statistics of anaphoric expressions as to topicality.

Anaphoric form	Topicality				
	antecedent topical N	%	antecedent non-topical N	%	total
ZA	203	93.1	15	6.9	218
Pronoun	78	87.6	11	12.4	89
demonstrative	0	0	1	100	1
dem + bare NP	4	33.3	8	66.7	12
bare NP	19	34.5	36	65.5	55
Poss NP	2	28.6	5	71.4	7
Proper noun	39	60.9	25	39.1	64
rel + NP	3	50	3	50	6
Poss NP + PN	1	50	1	50	2
total	349	76.8	105	23.2	454

Table 12 Breakdown of Anaphoric Expressions in Written Texts as to Topicality

Anaphoric form	Topicality				
	antecedent topical N	%	antecedent non-topical N	%	total
ZA	163	88.1	22	11.9	185
Pronoun	245	83.9	47	16.1	292
demonstrative	2	50	2	50	4
dem + bare NP	39	29.5	93	70.5	132
dem + PN	15	78.9	4	21.2	19
bare NP	10	25.6	29	74.4	39
Poss NP	26	44.8	32	55.2	58
Proper noun	49	49	51	51	100
rel + NP	1	14.3	6	85.7	7
Poss NP + PN	3	60	2	40	5
total	553	524.1	288	475.9	841

Table 13 Breakdown of Anaphoric Expressions in Spoken Texts as to Topicality

Table 12 and Table 13 show that in both spoken and written texts, ZA and pronouns mark their antecedents as predominantly (local) topics. Significantly, all other types of anaphors (with the exception of proper nouns and demonstrative plus proper noun), namely the lower accessibility markers, favor their antecedents as non-topics. This apparently is attributable to either the fact that they are more likely to refer to nonhumans (76.4% of bare NPs refer to nonhumans), or the fact that they often refer to entities of secondary or tertiary importance in the narratives (e.g. possessive NPs and determinate NPs).

Since topics are more accessible than non-topics, and since high accessibility markers refer to contextually salient (topical) entities, and low accessibility markers to less salient entities, such as non-topics, it should follow that high accessibility markers (ZA and pronouns) should be better retrievers for topics and low accessibility markers better retrievers for non-topics.

2.4. Accessibility and Unity

Another, and the last, factor affecting accessibility is unity, or textual cohesion of anaphors and antecedents. One way to measure unity is compute degree of topic continuity across sentences or topic chains. According to Ariel (1990: 29), unity depends on whether antecedents are within the same frame, same world, same point of view, same segment or paragraph as the anaphors. Although sentences, topic chains and paragraphs each constitute a coherent unit, they differ in degree of cohesion, since paragraphs are made up of sentences or topic chains which in turn are composed of clauses. The beginning of a discourse unit such as topic chain or paragraph is necessarily less cohesive with the preceding discourse unit than neighboring linguistic expressions within the same discourse unit. As an attempt to understand unity as a measure of accessibility, we compute (1) the number of various referring expressions at the beginning of sentences, topic chains and paragraphs and determine the preferred forms marking topics continuous with the preceding discourse unit; (2) the number of various referring expressions used to introduce intra-text or extra-text entities.

Table 14, Table 15, and Table 16 present the relevant statistics.

Anaphoric form	First anaphor of sentence, topic chain or paragraph			
	sentence (writing)	paragraph (writing)	topic chain (speech)	paragraph (speech)
ZA	22 (18.8)	1 (2.6)	0	0
Pronoun	46 (37.7)	2 (5.1)	44 (32.6)	9 (24.3)
determ NP	3 (2.5)	3 (7.7)	28 (20.7)	6 (16.2)
bare NP	15 (12.3)	2 (5.1)	6 (4.4)	1 (2.9)
Poss NP	6 (4.9)	0	16 (11.9)	3 (8.1)
Proper NP	18 (14.8)	29 (74.4)	29 (21.5)	10 (27)
Complex NP	5 (4.1)	1 (2.6)	3 (2.2)	3 (8.1)
* indef NP	7 (5.7)	1 (2.6)	9 (6.7)	5 (13.5)
total	122 (100)	39 (100)	135 (100)	37 (100)

Table 14 Breakdown of Anaphoric Expressions as Discourse Unit Initials in Written and Spoken Texts

Anaphoric form	First anaphor of sentence			First anaphor of topic chain		
	continuous	discontinuous	total	continuous	discontinuous	total
ZA	18 (81.8) (27.7)	4 (18.2)	22	0	0	0
Pronoun	42 (91.3) (64.6)	4 (8.7)	46	37 (84.1) (86)	7 (15.9)	44
determ NP	1 (33.3)	2 (66.7)	3	2 (7.1)	26 (92.9)	28
bare NP	1 (6.7)	14 (93.3)	15	1 (16.7)	5 (83.3)	6
Poss NP	0	6 (100)	6	0	16 (100)	16
Proper NP	3 (16.7)	15 (83.3)	18	3 (10.3)	26 (89.7)	29
Complex NP	0	5 (100)	5	0	3 (100)	3
* indef NP	0	7 (100)	7	0	9 (100)	9
total	65 (53.3)	57 (46.7)	122	43 (31.9)	92 (68.1)	135

Table 15 Breakdown of Anaphoric Expressions as to Topic Continuity in Written and Spoken Texts

There is an unmistakable tendency for the spoken texts, in comparison with the written texts, to use lower accessibility markers to begin a discourse unit, as can be seen in Table 14. Thus high accessibility markers (ZA and pronouns) account for 56.5% of all sentence-

introducers as against just 32.6% for topic chain-introducers. This comports with the earlier finding that antecedents for anaphoric expressions in speech consistently favor the more recent positions than those in the written texts, and that the preferred position for the antecedents of both ZA and pronouns in speech is the previous clause (see Table 8), which is a more recent position than the topic chain. This explains nicely why the spoken texts need more informative referring expressions to introduce topic chains.

ZA and pronouns have been shown to favor their antecedents as topics (Table 12 and Table 13). Furthermore, they account for an overwhelming proportion of all occurrences of anaphoric expressions marking topics continuous between sentences (92.3%) or between topic chains (86%), as can be seen in Table 15. In other words, ZA and pronouns tend to maintain continuous topics across sentences/ topic chains, while all the other expressions tend to signal discontinuous topics (determinate NPs, mid accessibility markers on other measures, failed to show an intermediate pattern of distribution on this count). This is not surprising since it is well known that there is a strong discourse pressure in Chinese, as in other languages, to mark initial NPs as topics in successive clauses/ sentences/ topic chains, under the reasonable assumption in human communication that all other things being equal, the current clausal topic also tends to be the topic in the next clause, a consequence of the principle of coding economy in the structure of memory and capacity of attention (Huang 1991).

Table 16 shows that the low accessibility markers are also the most popular markers for introducing extra-text entities: complex NPs, possessive NPs, bare NPs and proper nouns, in descending order of popularity, which is in close correspondence to the accessibility scale established earlier, whereas no single occurrence of ZA or pronoun serves that function. This is to be expected, since low accessibility markers are commonly used on the basis of general knowledge in unmarked initial referring acts and high accessibility markers are normally thought of as requiring a linguistic antecedent.

discourse type / anaphor form	writing			speech		
	intra-text	extra-text	total	intra-text	extra-text	total
ZA	218 (100)	0	218	185 (100)	0	185
Pronoun	89 (100)	0	89	292 (100)	0	292
demonstrative	1 (100)	0	1	4 (100)	0	4
dem + NP	12 (100)	0	12	132 (100)	0	132
dem + PN	0	0	0	19 (100)	0	19
bare NP	55 (79.7)	14 (20.3) (23..3)	69	39 (75)	13 (25) (27.6)	52
Poss NP	7 (63.6)	4 (36.4) (6.6)	11	58 (76.3)	18 (23.7) (38.3)	76
Proper noun	64 (80)	16 (20) (26.6)	80	100 (91.7)	9 (8.3) (19.1)	109
Complex NP	8 (53.3)	7 (46.7) (11.6)	15	12 (66.7)	6 (33.3) (12.7)	18
* indef. NP	0	18 (30)	18	0	31 (65.9)	31
total	454 (88.5)	60 (11.5)	513	841 (94.7)	47 (5.3)	888

Table 16 Breakdown of Anaphoric Expressions as to Endophoricity in Written and Spoken Texts

It is important to note that **all** of the occurrences of determinate NPs refer intra-textually, which is strong evidence that the demonstratives, both proximal and distal, have been grammaticized as endophoric markers akin to the definite article in English.

Since new entities are more often introduced into discourse in the object (or oblique object) position, and only rarely in the subject position (Huang 1991, Du Bois 1987), we would predict them not to occur as the first anaphor of a discourse unit and at the same time mark topics continuous with the preceding discourse unit. This prediction is borne out by Table 15: there complex NPs, possessive NPs, proper nouns and bare NPs are indeed the most popular expressions to mark discontinuous topics.

2.5. Summary

We now summarize the various accessibility scales of the anaphoric expressions based on the four factors of distance, competition, saliency and unity.

(1) Distance

    A. writing

        ϕ > pronoun > determinate NP > bare NP > possessive NP > proper noun >

        complex NP

    B. speech

        ϕ > pronoun > determinate NP > bare NP > proper noun > possessive NP >

        dem plus PN > complex NP

(2) Competition

    A. writing

        ϕ > pronoun > determinate NP > proper noun > bare NP > poss NP > complex NP

    B. speech

        ϕ > pronoun > determinate NP (excluding dem plus PN) > bare NP > poss NP >

        proper noun > dem plus PN > complex NP

(3) Saliency

    A. writing

        ϕ > pronoun > proper noun > complex NP > bare NP > determinate NP > poss NP

    B. speech

        ϕ > pronoun > dem plus PN > proper noun > poss NP > complex NP >

        determinate NP > bare NP

(4) Unity

    4.1. Topic Continuity

    A. writing

        ϕ > pronoun > determinate NP > proper noun > bare NP > poss NP

                                    complex NP

    B. speech

φ > pronoun > determinate NP > proper noun > bare NP > poss NP

                                                    complex NP

4.2. Intra-textuality

A. writing

ZA   > proper noun > bare NP > poss NP > complex NP

pronoun

det NP

B. speech

ZA   > determinate NP > proper noun > poss NP > bare NP > complex NP

pronoun

dem plus PN

That the same anaphoric expressions were placed differently on different scales should not be surprising, since accessibility is a multi-dimensional concept and each factor measures a different component of accessibility. Still, the general picture that emerges from the ten measurements is clear enough: in the written texts, ZA and pronouns are high accessibility markers, determinate NPs are mid accessibility markers and the rest low accessibility markers. Among the low accessibility markers, proper nouns are highest in accessibility, followed by bare NPs, possessive NPs and complex NPs, in descending order of accessibility.

The rank order among low accessibility markers in the spoken texts is much more complex, but can be determined by using a rating system that awards 5 points to first place on any scale, 4 points to second place, 3 points to third place, 2 points to fourth place and 1 point to fifth place. The resulting composite accessibility scale is exactly as in the written texts, except for the additional presence of dem plus PN:

proper noun > dem plus PN > bare NP > possessive NP > complex NP.

3. Degrees of Accessibility: Finer Distinctions

We have established in the previous sections a general accessibility scale of the anaphoric expressions in both written and spoken texts. But there is more to an anaphoric expression than its placement on the accessibility scale. The purpose of this section is to determine the raison d'etre of each expression in the referential system of the language as a whole in order to understand finer distinctions among the high, mid and low accessibility markers. We begin with the two high accessibility markers.

3.1. High Accessibility Markers: ZA and Pronoun

ZA has been shown in the earlier sections to be the best retriever for antecedents across clauses and pronouns the best retrievers for antecedents across sentences/ topic chains. This

45

division of labor between ZA and pronouns makes eminent linguistic sense since clauses and sentences (or topic chains) are different discourse units and since the linkage between clauses is tighter than that between sentences, it is only natural for languages to have developed a referential system in which the more attenuated form is used to achieve discourse cohesion between more tightly linked units and the fuller form used to achieve cohesion between more loosely linked discourse units (cf. Foley and Van Valin 1984). In processing terms, this means that when an entity is evoked by means of ZA or pronoun it is already highly activated in the hearer's focus set and is likely to remain so at the subsequent mention and no referential expression is needed as long as the intended referent is the one expected by the hearer. This observation, which formed the conerstone for Givon (1983)'s investigation into distance-based accessibility hierarchy across languages, has also been built into a number of processing systems as a pragmatic strategy in anaphor resolution, e.g. the focus strategy in Grosz (1982) and Sidner (1981), Levinson (1987)'s Q- and I-principles, and Chomsky (1981)'s "Avoid Pronoun" principle (cited in Ariel (1988)).

3.2. Mid Accessibility Markers: Determinate NPs

The status of determinate NPs appears to represent something of a mystery. Their seeming mystery can be seen by comparing dem plus bare NPs (i.e. bare NP modified by a demonstrative, distal or proximal), which constitutes the bulk of the superordinate category determinate NP (92.3% in writing and 85.2% in speech), with bare NPs, and dem plus PN (proper nouns modified by a demonstrative), also a subcategory of the determinate NP, with proper nouns with regard to their behavior on the various accessibility measurements.

We have established in the preceding section that in both written and spoken texts, dem plus bare NPs are consistently more accessible than bare NPs -- they were significantly more accessible in 4 out of 9 measurements; in the other 5 measurements, their differences did not reach significant level. In the spoken texts, proper nouns were more accessible than dem plus PN. In other words, a demonstrative in construction with bare NP makes the whole anaphoric expression more accessible, whereas when it is in construction with a proper noun, it makes the whole expression less accessible. Why should this happen?

On closer examination, proper nouns are found to be significantly more accessible than dem plus PN on both distance and competition measurements, but the reverse is the case on saliency and unity measurements. Why should this be the case? The reason seems to be this: all of the determinate NPs have been shown to be anaphoric in function, meaning that demonstratives have been grammaticized into quasi-definite article markers; if so, they are also more likely to be used than bare NPs or bare proper nouns when their antecedents are topics, other things being equal.

But while demonstratives are anaphoric in function, they are also semantically contentful, which explains why on both distance and competition measurements, dem plus PN is

consistently less accessible than unadorned proper nouns. But that still leaves unexplained the fact that determinate NPs (strictly, dem plus bare NP) are significantly more accessible than bare NPs. Given that demonstratives are lexically contentful elements, one would expect bare NPs to be the more accessible, just as unadorned proper nouns are more so than dem plus PN. Answers to this puzzle may well hinge on some critical difference between bare NPs and proper nouns, to which we now turn.

3.3. Low Accessibility Markers

Complex NPs are the lowest of the four low accessibility markers. This is understandable, since they consist of two subcategories (a combination of a relative clause and bare NP and a combination of a possessive NP and proper noun) and must therefore be less accessible than either bare NPs or proper nouns. Furthermore, possessive NPs are less accessible than bare NPs. This again is understandable, since they are formed by a combination of a possessive (personal pronoun or determinate NP) and bare NP and should therefore be less accessible than bare NPs. That leaves us with the question of why proper nouns should be more accessible than bare NPs, a fact already established in the earlier section.

The fact that proper nouns are more accessible than bare NPs, it turns out, turns critically on two measurements only: saliency and intra-textuality, as they are **not** significantly different from each other on all other measurements. Now on the saliency measurement, proper nouns are used to refer to topical antecedents significantly more often than bare NPs, clearly a consequence of the fact that proper nouns are more likely to refer to humans than do bare NPs (93% vs. 46.4% for the two texts combined). On the intra-textuality measurement, proper nouns are significantly more likely to refer to intra-textual entities than are bare NPs. This is so because bare NPs often refer to nonhumans, which are introduced into discourse chiefly in the object position where they are more likely to be forgotten, and not mentioned again.

Of course the more non-topical bare NPs may be rementioned, via the anaphoric determinate NP (in effect dem plus bare NP) in the immediately succeeding clause, forming an S=O anaphoric link across successive clauses, where S is subject of a following clause and O object of the preceding clause. In a count of anaphoric links across successive clauses according to the syntactic roles in which the coreferential mentions occur in the two clauses, I found that S=A links far outnumbered S=O links by a ratio of four to one (80% vs. 20%) (Huang 1991). Since proper nouns are more topical, hence more accessible, their second mentions are likely to be done via higher accessibility markers, reserving the more informative dem plus PN for more distant antecedents.

4. Conclusion

I hope to have shown that referring expressions in Mandarin Chinese are specialized as to the degree of accessibility they mark. Four types of distributional facts were used to establish the accessibility scale for anaphoric expressions in both written and spoken texts. The general accessibility scale we have arrived at is as follows:

ZA > pronoun > determinate NP > proper noun > bare NP > poss NP > complex NP

ZA and pronouns are high accessibility markers, determinate NPs mid accessibility markers and the rest low accessibility markers. When the speaker uses a high accessibility marker, he assumes its antecedent is the most accessible (i.e. activated) to the hearer, while use of a low accessibility marker means its antecedent is the least accessible.

Marker accessibility is closely tied to context types and structure types. In an initial referential act, the speaker commonly uses a low accessibility referring expression and the speaker refers the hearer to the most general knowledge for referent interpretation. By contrast, recent linguistic material is the most accessible information and can be accessed by the use of the semantically emptier high accessibility markers on a second mention.

High accessibility markers have the most recent antecedents and low accessibility markers the most remote ones. Specifically, zero anaphors are the best retrievers for antecedents across clauses and pronouns the best retrievers for antecedents across sentences. Among the four low accessibility markers we have been able to establish their relative rank order on the accessibility scale on the basis of ten measurements bearing on their behavior with regard to referential distance, competitiveness, saliency and unity. While not all measurements yielded the same accessibility scale, there were enough statistically significant differences between relevant measurements for us to deduce the final accessibility scale.

A major finding of this study has been that for each anaphoric expression, its antecedents in the spoken texts consistently favor the more recent positions than those in the written texts. This means that written texts should be preferred if our primary purpose is to find out the favored antecedent position for each anaphoric form, given the fact that there is general conflation among the favored positions in spoken texts. However, spoken texts do have the virtue of displaying a greater diversity of anaphoric expressions for us to work with, an important advantage if our purpose is to better understand the working of a referential system, such as the one undertaken in this study.

FOOTNOTES

1. I am immensely grateful to Lin I-ling for help with analysis of the data presented here. This paper draws heavily on her M.A. thesis (Lin 1992), but addresses a set of issues not central to her concern and arrives at some substantial different conclusions.

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

3. The significance of the length of a clause in the written texts being longer than that in the spoken texts does not escape me, but will not be addressed in this study. Cf. Chafe (1987b)

4. See Chafe (1987b) and Halliday (1987) for discussion of this issue.

REFERENCES

Akmajian, A. and Jackendoff, R. 1970. "Coreferentiality and Stress". Linguistic Inquiry 1, 124-126.

Anderson, John. 1976. Language, Memory and Thought. Lawrence Erlbaum.

Ariel, Mira. 1988. "Referring and Accessibility." Journal of Linguistics 24: 65-87.

_____. 1990. Accessing Noun Phrase Antecedents. New York: Routledge.

Bolinger, D. 1977. Pronouns and Repeated Nouns. Indiana University Linguistics Club.

_____. 1979. "Pronouns in Discourse." In T. Givon (ed.) Syntax and Semantics 12. New York: Academic Press.

Chafe, Wallace. 1980. "The Deployment of Consciousness in the Production of a Narrative." In W. Chafe (ed.) The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production, 9-49.

_____. 1987a. "Cognitive Constraints on Information Flow." In Tomlin, R. (ed.) Coherence and Grounding in Discourse, 21-51.

_____. 1987b. "The Relation between Written and Spoken Language." Annual Review of Anthropology 16: 383-407.

Charniak, E. 1972. "Toward a Model of Children's Story Comprehension." AI TR-266, MIT AI Lab.

Chomsky, N. 1981. Lectures on Government and Binding. Dordrecht: Foris.

Collins, A. and Quillian, M. 1972. "How to Make a Language User." In E. Tulving and W. Donalson (eds.) Organization of Memory, 109-351. Academic Press.

Du Bois, J. W. 1987. "The Discourse Basis of Ergativity." Language 63.4.805-55.

Foley, W. A. and Van Valin, R. D. 1984. Functional Syntax and Universal Grammar.Cambridge University Press.

Givon, T. 1983. (ed.). Topic Continuity in Discourse: a Quantitative Cross-Language Study. Armsterdam: Benjamins.

Grosz, B. J. 1981. "Focusing and Description in Natural Language Dialogues." In A.Joshi (ed.) Elements of Discourse Understanding. 84-105. Cambridge University Press.

Halliday, M. A. K. 1987. "Spoken and Written Modes of Meaning." In R. Horowitz et. al. (eds.) Comprehending Oral and Written Language, 55-81. Academic Press.

Huang, Shuanfan. 1991. "Chinese as a Discourse Accusative Language." paper presented at the Third International Conference on Chinese Language Teaching and Research. Taipei,Taiwan.

Keenan, E. 1976. Anaphora and Cross Referencing Systems. Indiana University Linguistics Club.

Levinson, S. C. 1987. "Pragmatics and the Grammar of Anaphora." JL 23.379-434.

Li, C. and S. Thompson 1979. "Third-person Pronouns and Zero Anaphora in Chinese Discourse." In Givon (ed.) Syntax and Semantics 12.

Lin, I-ling. 1992. Anaphora in Written and Spoken Chinese Narratives. M.A. thesis, National Taiwan Normal University.

Sidner. C. L. 1982. "Focusing for Interpretation of Pronouns." Am. J. of Computational Linguistics 7:4, 217-231.