

A Comparison of Learner and Native Speaker NP Ellipsis in Mandarin

Helen Charters

Introduction

English may be said to be a "hot" language' because pronouns cannot in general be omitted from grammatical sentences, and the information required to understand each sentence is largely obtainable from what is overtly seen or heard in it. On the other hand, Chinese may be said to be a very "cool" language in that such pronouns are usually omissible (and are often more naturally omitted). - C. T. James Huang (1984)

This article focuses on the question of the frequency of ellipsis in informal written Mandarin texts. It compares the output of native speakers with that of learners in their second year of study, describing their use of NP ellipsis, pronouns, and nouns in short written texts. Subsequent analysis (see Charters 1996a) investigates the syntactic structures in which ellipsis occurs.

It is hoped that a better understanding of such differences will suggest areas where explicit instruction might be profitably given, and indicate areas where it is unnecessary.

Ellipsis in Mandarin

Modern Standard Chinese (Mandarin) has regular definite ellipsis, that is, any NP can be ellipsed so long as its referent is retrievable from context (Fillmore, 1986; Shopen, 1972, 1973). It also has lexical indefinite ellipsis, that is, some verbs have intransitive as well as transitive valency options. Intransitive counterparts of a normally transitive verb, like the use of 'kill' in 'he has killed twice', represent a distinct phenomenon, both syntactically and referentially, from the omission of arguments with a definite, specific and retrievable interpretation. It is the latter which account for the greatest differences between English and Mandarin referential strategies.

Mandarin has no verbal agreement or other explicit co-indexing strategies. The minimal sentence in Mandarin is a verb with no realised arguments at all:

¹ Marshall McLuhan (1964) characterises a "hot" language as one involving little audience participation and a "cold" one as one requiring active audience participation.

- 1) (Ø) yào (Ø)
 (Ø) want (Ø) (Ø = predictable entity i.e. definite ellipsis).

Acquisition of zero anaphora

Most studies of learners' use of Ø anaphora have involved English, where NP ellipsis is relatively constrained, as the target language (see for example White, 1985, Williams 1989). These studies have found a greater use of zero anaphora by learners than is customary for native speakers, even using descriptive, rather than prescriptive, standards of comparison (White 1985, Williams 1989).

Three explanations have been offered for the differences in employment of anaphoric devices in learner Inter-Language: firstly that ellipsis is a device employed by learners to reduce the demands on their production, when the referent is readily available in the discourse (Klein 1986, Williams 1989); secondly that they are the result of L1 interference (White 1985); and thirdly that they may result from universal tendencies (Gundel and Tarone, 1983).

For native speakers of English, where ellipsis is highly constrained, learning Mandarin, where definite ellipsis is free, the effects of 'limited resources' vs transfer will be in contrast: if learner ellipsis is largely the consequence of 'limited linguistic resources', we can expect English-speaking learners of Mandarin to also display a relatively high degree of ellipsis. On the other hand, if it is largely the consequence of L1 transfer, we can expect learners to display levels of ellipsis considerably less than those of native speakers. Thus a study of the acquisition of ellipsis in Mandarin by English-speaking learners is a useful contribution to an inquiry into L1 transfer.

Of course the patterns of acquisition and use arising out of L1 transfer may be complex. Larsen-Freeman and Long (1991) state that while the Contrastive Analysis Hypothesis predicts that L1 transfer may either cause difficulty or facilitate acquisition of L2, 'recent research, however, has shown that transfer manifests itself in unexpected ways as well' (Larsen-Freeman and Long, 1991:97). Specifically, L1 transfer can prolong the period of error commission in areas of typological contrast between L1 and L2; '[z]ero contrast (i.e. the L2 possesses a category that is absent in the learner's L1) affects I[nter] L[anguage] in more subtle ways than originally believed. One effect is to delay passage through a developmental sequence...' (Larsen-Freeman and Long, 1991:99). Such a typological contrast pertains for Mandarin and English in the area of ellipsis: Mandarin has regular definite ellipsis, while English does not. Alternatively, L1 transfer may result in overuse of a parallel L2 structure (Larsen-Freeman and Long, 1991:99). This may account for the overuse of ellipsis in English by Chinese and Spanish-speaking learners, as discussed below.

Eckman's (1977, 1985) Markedness Differential Hypothesis predicts

A Comparison of Learner and Native Speaker NP Ellipsis in Mandarin

that those areas of the L2 which differ from the L1, and are more marked than L1 will be difficult to acquire. On the other hand, there is some evidence of elementary learners transferring more marked forms from L1 into L2 (Larsen-Freeman and Long, 1991).

Each of these predictions has some relevance to the issue of contrasting patterns of ellipsis in Mandarin and English and they will be taken up again in the discussion below.

Past research

There is an extensive literature on anaphora in general, and more particularly on zero anaphora, but there is rather less dealing specifically with Mandarin. Fillmore (1986) and Shopen (1972, 1973) discuss a typology of ellipsis with reference to English discourse and semantics; Aoun (1985), and Givon (1983, 1984), discuss the use of different anaphoric devices, including overt and covert forms, in referent tracking; Chomsky (1980, 1981, 1982), Chung (1989), Farmer and Harnish (1987), Haegeman (1991), Kaplan and Bresnan (1982) Levinson (1983, 1987a, 1987b, 1991), Reinhart (1976, 1983) and Rizzi (1986), among others, discuss the distribution and interpretation of zero anaphors and pronouns, as well as other anaphoric and syntactic devices; Lust (1986) focuses on the question of the acquisition of anaphora in general, not just zero anaphora; James Huang (1982, 1984, 1989, 1993), Huang Yan (1992, 1994), Li and Thompson (1979), Chen Ping (1984), and Battistella (1985) all discuss zero anaphora with specific reference to Mandarin, though they take very different approaches to the subject.

Learner use of zero anaphora

Givon (1983) proposes that all speakers conform to a suggested universal hierarchy of topic accessibility, in that they use more marked forms (ie more morphologically and phonologically complex) when referring to less accessible topics and vice versa. Givon (1984) compares transcripts of spoken narrative from two speakers of Hawaiian Pidgin English, one Korean and one Filipino, and conversation of a Spanish speaker of 'early stage English'. He uses referential distance as the measure of discourse function. Referential distance is quantified as the average of the number of clauses between one mention of an entity and the nearest previous mention of the same entity (Givon, 1983:13). Distances greater than 20 clauses are deemed to constitute re-establishment of a topic, and not continuity of reference.

In Givon's 1984 study, the Korean-English text has more zero anaphora than pronouns in both subject and object role, and regardless of person. By contrast, the Filipino has more pronouns than zero anaphora in all cases, though the figures for Object pronouns are so low as to be inconclusive. The Korean speaker's usage also differs from that of the Filipino in terms of referential distance. The Korean's average referential distance for pronouns is higher in 1st and 2nd person (mean referential

distance of 3.46 clauses). Givon puts this down to the Korean's greater use of zero anaphora at short distances. The relative lack of 1st and 2nd person overt pronouns at short distances raises the Korean's overall mean referential distance for pronouns.

In 3rd person references, the Korean subject's mean value for both zero anaphora and pronouns is the minimum: 1 clause. This is consistent with the fact that 1st and 2nd persons are more contextually immediate than 3rd persons, i.e. a less marked context. The Korean's use of nouns for 3rd person references is therefore consistent with Givon's predicted correlation.

On the other hand, in 3rd person, the Filipino subject's referential distance for pronouns is much higher than that of the Korean. Since the Filipino subject uses primarily pronouns even at short distances, avoiding zero anaphora, the Filipino's high mean does not reflect a differentiation between zero for short distances and a pronoun for longer ones. It must actually reflect a greater range of values for pronouns, including some much longer distances. This illustrates that the choice of referential device for a specific referential distance varies with L1.

Givon (1984) also suggests that learners' deviations from the values for specific devices expected of 1st-language users, are typically instances of using more marked (bigger, more morphologically complex, more semantically rich etc) forms at 'a much less marked functional point' (i.e. in environments of much higher topic-continuity).

However, there is some evidence from his 1984 study that this prediction does not hold true for the part of the continuum covered by zero anaphora and pronominals. Both Givon's (1984) Korean-English text, with a mean referential distance of 3.46 for 1st and 2nd person Subject pronouns, and the Philippine-English text, with a mean referential distance of 3.6 for 3rd person Subject pronouns, surpass the mean referential distance cited by Givon (1983, 1984) for pronoun use in Native English texts. He states (1984:131) that in English, pronouns have a mean referential distance of 1, the same as for zero anaphora. In this case native speakers must be using nouns (more marked) for greater referential distances, where the Korean and Filipino were using pronouns.

Williams (1989) and others also report considerable use of ellipsis (less marked) by learners of English where native speakers would use pronouns (more marked). Again this is contrary to Givon's (1984) prediction. Givon suggests the overlap of scores for pronouns and zero anaphora is an indication that the two devices serve much the same discourse functions, both within and across languages. If we accept this, then we would expect the variation to be at the level of chance, but the data for Mandarin, presented below, show significant differences in the frequency of ellipsis used by native speakers and learners of Mandarin.

Results of Williams' (1989) study suggest a more complex picture. Williams (1989) compares the functions of pronoun copies, pronouns and zero anaphora, in varieties of spoken English: Singapore English, American

A Comparison of Learner and Native Speaker NP Ellipsis in Mandarin

English, and the English of learners from a variety of language backgrounds. She uses the measures of referential distance, persistence and ambiguity, as developed by Givon (1984) to establish the discourse function of each device.

Williams (1989:185) draws the conclusion that there are no significant differences between groups in terms of the referential distances over which they use each anaphoric device. In fact though, her data show that for native speakers of both varieties of English, 'zero anaphora' has a much shorter referential range than pronouns (1.88 vs 3.46 for American English, and 3.19 vs 4.24 for Singapore English) whereas for the ESL subjects, there is no significant difference between the referential distances of 'zero anaphora' and pronominals (3.52 and 3.12 respectively). Despite Williams' conclusion, the figures suggest that learners are using zero anaphora in contexts where native speakers use pronouns, contrary to Givon's prediction. This difference is substantially reduced for speakers of Singapore English. Williams suggests (1989) the greater range of zero anaphora in the varieties of both Singapore English speakers and learners may result from L1 transfer.

There are also highly significant differences in Williams' study, in the frequency with which different groups use zero anaphora in specific contexts. This is true for both discourse contexts (introducing, reinstating and continuing a topic) and syntactic contexts (coordinate and non-coordinate structures). Native speakers of American English restrict their use of zero anaphora to co-ordinate clauses, or non-co-ordinate clauses where the ellipsed argument and its antecedent have the same Grammatical Function. Singapore English speakers and learners were found to use zero anaphora (ellipsis) in *non-co-ordinate* contexts, a departure from standard English norms, and to use it *more frequently* in grammatically acceptable contexts.

Despite these observations, and the fact that 'the functions of the other devices are not as clearly delineated as that of pronoun copies' (Williams 1989:182), Williams concludes that the functions of all the linguistic devices studied 'remains relatively constant.' (1989:185). She therefore accounts for the observed variability in the use of anaphoric devices, not on functional grounds but on the grounds of the pressures of unplanned production and immediate context, both for native speakers, and more so for learners.

I suggest that Williams' conclusion is incorrect, and that \emptyset anaphora may have different discourse functions from overt pronouns in these varieties of English.

White (1985) investigates the effect of L1 on English acquisition, comparing learners who are Native speakers of Spanish, a language with NP ellipsis, and French, without NP ellipsis. She found that L1 affected learner judgements on the grammaticality of ellipsed Subjects, and judgements improved with proficiency, but there was considerable variability between specific sentences.

One account of learning in the opposite direction, i.e. from a constrained ellipsis to a relatively free ellipsis language, is that of Fakhri, (1984). Fakhri looked at an English speaker learning Moroccan Arabic, where Subject pronouns can be ellipsed, but verb agreement provides referential information. This research found that the learner used full pronouns to compensate for the lack of verb-agreement (which she had not fully mastered), where (she considered) her verb agreement was accurate, pronoun use was reduced. This suggests a sensitivity on the part of the learner, to the anaphoric roles of both verbal agreement and pronouns, and a reluctance to omit pronouns unless the information they convey is signalled elsewhere.

Ellipsis in Mandarin

The majority of the works on zero anaphora in Mandarin have been primarily concerned with the processes by which interpretation is achieved, some claiming that co-reference is free (Li and Thompson, 1979; La Polla, 1993, 1994) some that it is syntactically constrained, (James Huang, 1982, 1984, 1989; Battistella, 1985) and some that it is pragmatically determined (Huang Yan, 1992, 1994).

According to Li and Thompson (1979), and La Polla (1993, 1994) the interpretation of ellipsed NPs in Mandarin is pragmatically determined and not predictable from the grammatical role of the omitted NP or its co-referent. By contrast, James Huang (1984, 1989) adopts and modifies the standard position within a Government and Binding theoretical framework, that 'empty constituents' in Mandarin are syntactically bound or controlled. James Huang (1984, 1989) suggests that ellipsed NPs have different constraints upon their interpretation than lexicalised NPs, and moreover that the contrast between overt and ellipsed NPs patterns differently in Mandarin from the same contrast in English.

For example the English sentence:

2) 'John says Tony doesn't know him'

is ambiguous as to whether the pronoun 'him' refers to 'John' or to some other participant in the discourse; the English sentence

3) *John says Tony doesn't know \emptyset '

cannot refer to a human participant at all, be it John, or some other participant, though it may refer to an abstract 'piece of knowledge' so long as the referent is already established in the discourse.

Similarly, the Mandarin sentence

A Comparison of Learner and Native Speaker NP Ellipsis in Mandarin

4) Zhāngsān shuō Lǐsìbù rènshi tā

Z say L NEGrecognise3sg
'Zhangsan says Lisi doesn't know him/her.'

is also ambiguous between reference to the matrix Subject or to another entity. However, according to James Huang (1984, 1989) the Mandarin sentence

5) Zhāngsān shuō Lǐsìbù rènshi Ø

Z say L NEGrecognise
'Zhangsan says Lisi doesn't recognise him/her/it.'

can only refer to someone or something **other** than Zhangsan, unlike the English 3). The actual reference of this sentence is, in fact, somewhat controversial (see Huang Yan, 1994 for a refutation of James Huang's position). Asked to comment on such sentences in context, native speakers demonstrate wide variation in judgements of acceptability and reference (see Charters 1996b:80-82). In any event, a sentence of the structure of 5) would be ungrammatical in English with a concrete referent.

The question of the distribution of ellipsis is addressed only incidentally by James Huang (1984, 1989) and Huang Yan (1992, 1994) in that they organise their discussions of interpretation around instances of ellipsis in specific syntactic frames. James Huang gives particular attention to the reference of Ø Objects in sentential complements, as just discussed, and also in Topic-comment structures. Huang Yan discusses eight different syntactic structures in some depth, including relative clauses, correlative constructions and both Subjects and Objects of sentential complements. The only study where distribution is the main focus is Chen Ping's (1984). This however starts from the assumption that ellipsis is the norm, and it is the distribution of overt pronouns that requires explanation.

None of these works addresses the question of the frequency of ellipsis, nor its use by non-native speakers, which are the concerns of this paper.

Differences in the distribution of ellipsis in English and Mandarin

The only NP ellipsis generally considered grammatical in isolated English sentences is that of the Subject NP in co-ordinate clauses (Quirk et al., 1985). Williams (1989) presents evidence that spoken English actually allows more wide-spread Subject ellipsis in coordinate structures. However this is still more restricted than the range of Chinese ellipsis.

Definite ellipsis is regular (ie not lexically restricted) in Mandarin, while it is a-typical in English, and in Object position, virtually non-existent.

Givon (1984) suggests that the use of anaphoric devices follows universal discourse principles, whereby the least marked anaphoric devices refer to the most accessible referents. At the same time though, formally equivalent anaphoric devices may function over different parts of the topicality hierarchy in different languages (Givon 1984), so learners of a second language may employ specific anaphoric devices in patterns different from those of native speakers.

Thus we see that ellipsis may have a different distribution, serve different functions, and yield different interpretations in English and Chinese. Despite this, discussion of anaphoric patterns is not included in any of the main textbooks used to teach Chinese in Australia or New Zealand; students of Mandarin tend to be given little or no instruction, in the first few years, on writing texts longer than a sentence, much less explicit instruction regarding NP ellipsis. The study discussed in the remainder of this article represents a first step in gaining a clearer understanding of how learner and native patterns of ellipsis differ in Mandarin.

The current study

Methodology

SUBJECTS

Written texts were collected from eight native speakers of Mandarin aged between 20 and 50. There were four women and four men, all are residents of the PRC and seven have never been outside Mainland China. They live or grew up in areas where the local dialect is Mandarin (Beijing, Henan, Hubei, Dongbei). Five of the native-speakers were university educated, the others have graduated from high school. Two speak English well enough to study at English-speaking universities, the others have, practically speaking, little or no English.

Non-native subjects were students enrolled in second and third year Chinese classes at The Australian National University, Canberra, and Massey University, Palmerston North, NZ. There were eight Australian students, four men, four women, and six NZ students. All were aged between 20-40, except for one who was under 20.

All English speakers are native English speakers. Three of the subjects have a bi-lingual back-ground, but in neither case was their non-English language Chinese. In two cases the bilingual subject was educated entirely in English, and the other was educated in English after primary school level. The subjects' previous exposure to Chinese was somewhat mixed: three had lived in China for more than a year, and one for four months. Another had spent two and a half years in a Cantonese-speaking environment. However, despite the variation in time spent in previous study, seven of the Australian students, and all the NZ students were studying at the same level (year two), indicating that they have similar levels of proficiency. Available subjects were too few to allow for analysis of different proficiency levels.

The main hypothesis in this investigation is that L1 affects the

A Comparison of Learner and Native Speaker NP Ellipsis in Mandarin

acquisition of L2 patterns of zero anaphora. For English-speakers learning Chinese, learner texts are expected to have a greater frequency of lexical NPs, in both Object and Subject position, than native texts.

TEXTS

The native texts were extracts from letters to friends and relatives in Australia². Each text consists of at least 10 sentences of continuous text. In all cases the extract began at the point where the writer begins to introduce a new anecdote or piece of news. The learner texts were elicited in the second half of the school year via the subjects' usual class teachers. They were asked to write a connected paragraph of at least ten sentences. Where texts were elicited, subjects were asked to write on any topic, as if writing to a friend, though one learner text was part of a homework assignment on current affairs. Some topics were suggested, based on the content of the native speakers' letters, to maximise comparability of subject matter. Learner texts were shorter than native texts and were used intact. None of the subjects knew what the focus of the study was.

LINGUISTIC ANALYSIS

For each text the number of lexicalised and ellipsed Subject, Object and (for Copula constructions) Complement NPs was counted. Lexicalised NPs were divided into pro-nominal, nominal and DE Phrases.

The DE Phrase is a nominal modifier (relative clause) construction. The head of this construction can be ellipsed, in which case the DE Phrase is often considered to have a pronominal function. This is semantically and syntactically similar to English constructions using non-specific demonstrative pronouns:

- 6) bù néng zài xǐwǎnjīlǐ xǐ de Ø
 NEG can in dishwasher wash DE Ø
'Those Ø which can't be washed in the dishwasher...'

Relative clauses are generally thought to have a special discourse function: identifying a specific referent out of a known set, or introducing and identifying a new participant in the discourse. DE phrases were of very low frequency in the data, and since they serve different discourse functions from bare NPs, they were counted separately.

² My thanks to all those who participated in the project by writing texts specially, and also to Chen Yong and Chen Chun for generously sharing their personal correspondence with me. Thanks also to their friends and mine, the original authors of the texts. Where content is reported, names have been altered. Thanks also to Li Dong and Yang Tiejun for their co-operation in eliciting texts from their students.

The number of possible arguments in a text was assessed by considering the maximum number of arguments a specific verb could take, without introductory prepositions, **in the sentence structure as it occurred in the text.**

This means that the same verb may be assessed in different sentences as having a) different valency options; b) different expressions of the same underlying valency when constrained in a given construction.

Williams (1989) makes a useful distinction between NP omission, which is ungrammatical, and functionally unspecified, and zero anaphora, which may or may not be ungrammatical, and has a specified function i.e. co-reference. In analysing Chinese texts, it became clear that finer distinctions between types of ellipsis were necessary.

A three-way distinction is made between obligatory, contrastive and optional ellipsis.

For the purposes of this study, obligatory ellipsis includes cases where no NP can occur grammatically for instance between modal or auxiliary verbs and a main verb. Contrastive ellipsis refers to cases where an NP can occur but there are constraints on its interpretations, for example, where an NP co-referent with the Subject of another clause must be omitted, but a switch-reference must be lexicalised.

Obligatory ellipsis was not counted. Instances of optional ellipsis with no semantic change, or with epistemic changes, rather than referential changes were counted as ellipsed.

To illustrate consider:

Optional:

7) Shāndōng yǒu ge lǎorén yòng tā jiéshěngde qián mǎi diànshìjī

Shandong has CL old person use he save-DE money buy TV.
In Shandong there was an old man who used the money he'd saved to buy a TV.
(ns2:2-4)³

In this sentence the addition of independent NPs for each argument position breaks the structure up, but does not alter the meaning.

8) Shāndōng yǒu ge lǎorén yòng. Ta yòng tājiéshěngde qián.

Shandong has CL old person he use he save-DE money
In Shandong there was an old man. He spent the money he'd saved.

³ Reference numbers refer to lines in the collected texts: es 1-14 for English speakers; ns 1-9 for native speakers.

Tā mǎile diànshìjī

3sg buy-ASP TV
He bought a TV

This sentence was counted as having two ellipsed NPs.

The following example (from a non-native speaker) shows the use of verb serialisation to signal purpose.

9) Tā zhù zài xīnī gōngzuò.

3sg lives in Sydney work
He lives in Sydney [for his] work.

(es5:7)

If a lexical Subject NP is added to the second verb, the two predications appear to be unrelated and in this case, the second predication *tā gōngzuò* 'he works' is pragmatically very odd. Since the fact of 'working' is not especially newsworthy in this context, the second predication has neither a syntactic nor pragmatic link to the preceding discourse. It lacks a pragmatic nucleus (Liddicoat personal communication) and therefore seems incomplete.

10) Tā zhù zài xīnī. Tā gōngzuò.

3sg lives in Sydney. 3sg works.
He lives in Sydney. He works.

Ellipsis of the NP clearly serves to signal a relationship between the events. These instances were counted as ellipsis because it is precisely this kind of semantic and pragmatic control of omission as a device, which is relevant to the question of acquisition of native-like language use.

Contrastive:

11) Wǒ_i yào (*NP_i) bāngzhù tāmen zuò jiāwù.

I require (*NP_i) help (them) do housework.

I must help (them) do the housework.

(ns3:14)

In this example, an NP can occur as Object of the first verb and, simultaneously controls the reference of the Subject of the second; this Object, being of disjoint reference with the matrix Subject cannot be ellipsed. Conversely a Subject of the second verb which is co-referent with the Subject of the first may not be lexicalised.

12) Wǒ yào *wǒ/nǐ bāngzhù tāmen zuò jiāwù.

I require *1sg/2sg help (them) do housework.
I want *me/ you to help (them) do the housework.

(Interestingly the semantic role of the first NP changes from a kind of patient (obliged party) to a kind of agent (the requirer), but the syntactic status is the same.) Examples such as this involve syntactic processes of control that vary little from Mandarin to English. They are not contexts where pragmatic choices are available to speakers, and were excluded from this study (see Charters 1996b for discussion of their interpretation in native texts).

A rule of thumb was: if an NP could be added without affecting the grammaticality, and the action expressed by the verb was to be understood as affecting an entity represented by a specific antecedent (or post-cedent⁴) present in the discourse, or the writer or reader, (rather than a generic, non-referential 'referent'), then the sentence was assessed as having an ellipsed argument. When addition of a lexical NP altered the syntactic structure of a sentence, e.g. creating two independent sentences instead of a conjoint or sub-ordinate structure, but no significant meaning change accompanied this structural change, then this was also counted as ellipsis.

Because the relationships expressed via ellipsis are varied, the effect of lexicalisation on interpretation is also varied and the division between referential changes and epistemic changes is not completely clear-cut. Throughout the rest of this paper the term ellipsis will be used in the most general sense, to cover absence of any argument whether grammatical or ungrammatical, and whether or not presence and absence alters the interpretation. **Omission** will be used in Williams' sense of any **ungrammatical ellipsis**. **Optional ellipsis** refers to cases where the presence or absence of a lexical argument does not greatly affect the semantic interpretation. The term **zero anaphora** will be reserved for constrained ellipsis which affects the interpretation of referentiality. By these definitions, what is counted in this study is not zero anaphora, but optional ellipsis. However the structures involved are among those referred to as zero anaphora in Williams and Givon's studies.

Statistical Analysis

The data were analysed using χ^2 to determine whether there were significant differences between the frequencies with which native speakers and learners use ellipsis, pronominals and nouns in Subject and Object roles, and whether

⁴ In the case of embedded clauses (e.g. relatives) the Head usually follows the modifier, so that there is 'zero cataphora'. However, no analytical distinction was made on these grounds.

A Comparison of Learner and Native Speaker NP Ellipsis in Mandarin

there were within-group differences in the use of each device for Subject and Object arguments. Data for Complements and DE constructions were collected too, but the numbers were too small to be included in the χ^2 tests.

In addition, a three-way comparison was made between the data for native and learner Chinese, with Givon's (1984) data for a Korean speaker of Pidgin English. Although this data is from only one subject, it is based on a much larger text (7 pages) than those in the present study, so that the total possible arguments in each text type is roughly comparable. Since Korean allows free Subject ellipsis (Givon 1984) and it had been suggested that this text exhibited L1 influence, it was anticipated that the frequencies for the Korean-English text would be more similar to those of native Chinese than to learner Chinese. That is, that the effect of L1 influence would override the norms of the target language in both cases.

Results

The mean frequencies of each device are reported in Table 1. Native speakers were found to use ellipsis to a much greater extent than learners, though Object ellipsis is much less common than Subject ellipsis. Complement ellipsis is actually ungrammatical, and occurred only once in one learner text.⁵

Table 1. Summary Statistics: Mean Frequencies of Anaphoric Forms.

Form	Learner		Native	
	Means	sd	Mean	sd
Ø Subject	2.143	2.674	11.66	1.5
Realised Subject	12.286	3.604	12.22	5.142
Ø Object	(1 only)		1.889	1.364
Realised Object	9.714	5.967	12.444	4.876
Ø Comp	(1 only)		0	0
Real. Comp.	2.429	2.277	1.111	1.453
Sum of Ø NP	3.286	2.998	13.556	5.102
Sum Realised NP	23.786	9.601	25.444	8.338

⁵ This is most likely an attempt to use the DE construction, a nominal modifier with omitted Head, but the nominal marker DE was omitted also, resulting in an 'ungrammatical' construction. In fact a similar instance was found in a Native text, but not in Complement position.

Table 2 presents frequencies for the anaphoric devices for each Grammatical Role, with a breakdown of realised NPs into Pronouns, Nouns and DE constructions.

Table 2. Frequencies of Anaphoric Devices for Learners and Native Speakers

Form	Learner			Native		
	SUBJ	OBJ	COMP	SUBJ	OBJ	COMP
Ø	30	1	1	105	17	Ø
Pro	98	7	Ø	51	10	Ø
Noun	76	129	30	56	99	9
DE	Ø	Ø	4	3	3	3

One point that can be seen here is that learners use the DE construction in a more restricted context than Native speakers do, i.e. in Complement position only. (This is possibly the result of being taught a specific complement construction: *shi* *de* ' COP[modifier] DE' features in many text books.) Unfortunately the overall frequency of this construction in the texts is so low as to prevent statistical analysis. Studies of larger, or more texts may enable better analysis of its function. The overwhelming use of Subject ellipsis by native speakers is also evident.

Learner and native speaker usage of anaphoric devices was compared for Subject and Object functions separately, and χ^2 was used to test for statistical significance. Table 3 shows a significant difference ($\chi^2 = 59.39$, $df = 2$, $p < .0000001$) between native speakers' and learners' use of anaphoric devices in Subject function: native speakers use more ellipsis than expected, while learners use less. Pronouns group with Nouns, rather than with ellipsis. That is both lexical forms are used more than expected by learners, and less than expected by native speakers. The differences in Object position are also significant, though the overall frequencies are lower, and so is the level of significance ($p = .0001$, see Table 4).

By combining all Grammatical Roles, it was possible to get enough examples of the DE construction to include in a comparison. The observed values for DE structures differ only slightly from the expected values (Table 5), contributing very little to the overall differences.

The biggest deviation from chance patterns are in the use of ellipsis and full nouns. It is clear that it is primarily in the use of ellipsis that learner and native anaphoric strategies differ.

In order to investigate L1 transfer, Givon's (1984) data on Korean-English were compared to those of learner and native Chinese. Unfortunately, the Korean used no Object Pronouns and few Ø Objects so expected frequencies for Objects were too low for a χ^2 . Table 6 shows the results of a comparison for Subjects only. This Korean speaker of English

A Comparison of Learner and Native Speaker NP Ellipsis in Mandarin

Table 3

**Subject Anaphora:
Learners vs Natives**

	<i>Learner</i>	<i>Native</i>
∅	30 (66.20)	105 (68.8)
Pro	98 (73.07)	51 (75.93)
Noun	76 (64.73)	56 (67.27)

Table 4

**Object Anaphora:
Learners vs Natives**

	<i>Learner</i>	<i>Native</i>
∅	1 (9.38)	17 (8.62)
Pro	7 (8.86)	10 (8.14)
Noun	129 (118.77)	99 (109.23)

$\chi^2 = 59.39$ $p < .0000001$

$\chi^2 = 18.27$ $p = .00010778$

Table 5. All Core Arguments: Learner vs Native

<i>Form</i>	<i>Learner</i>	<i>Native</i>
∅	32 (78.48)	122 (74.51)
Pro	105 (85.16)	61 (80.84)
Noun	235 (204.68)	164 (194.31)
DE	4 (6.67)	9 (6.33)

$\chi^2 = 79.90$

$p < .0000001$

was found to pattern with the native Chinese rather than the English-speaking learners of Chinese, as predicted. In other words, the Korean's use of anaphoric devices was more similar to Chinese (and Korean) patterns even when speaking in English. (Although this is a comparison of spoken and written texts, Givon (1984) maintains that the discourse universals governing the choice of anaphoric device hold true across genres as well as languages).

Table 6. Subject Anaphora in Native and Learner Chinese compared with Korean-English

<i>Form</i>	<i>Learner Chinese</i>	<i>Native Chinese</i>	<i>Korean English</i>
Ø	30 (80.0z)	105 (83.23)	88 (59.68)
Pro	98 (67.88)	51 (70.54)	40 (50.58)
Noun	76 (56.03)	56 (58.23)	24 (41.75)

$$\chi^2 = 86.204 \quad p = .0001$$

FORM AND GRAMMATICAL FUNCTION

χ^2 was also used to determine whether there were within-group differences in the choice of anaphoric device for different Grammatical Functions (Subject and Object).

Table 7 presents the results for learners, native speakers, and Givon's (1984) Korean data.

Table 7. Form and Grammatical Function

<i>Form</i>	<i>Learners</i>		<i>Native Speakers</i>		<i>Korean</i>	
	<i>Subject</i>	<i>Object</i>	<i>Subject</i>	<i>Object</i>	<i>Subject</i>	<i>Object</i>
Ø	30 (18.55)	1 (12.45)	105 (76.52)	17 (45.48)	88 (74.4)	5 (18.6)
Pro	98 (62.81)	7 (42.18)	51 (38.26)	10 (22.74)	40 (32)	0 (8)
Noun	76 (122.64)	129 (82.36)	56 (97.22)	99 (57.78)	24 (45.6)	33 (11.4)

$$\chi^2 = 110.81 \quad p < .0000001 \quad \chi^2 = 86.69 \quad p < .000001 \quad \chi^2 = 73.588 \quad p = .0001$$

Despite the greater use of ellipsis by native speakers, both natives and learners use ellipsis predominantly with Subjects. Though native speakers use significantly more ellipsed Objects than learners, their frequency is still well below the expected values if the choice of device were purely a matter of chance. For all speakers there are significant differences in the frequency of each anaphoric device in each Grammatical Function. Subjects are consistently expressed by the less marked forms i.e. ellipsis and pronouns,

A Comparison of Learner and Native Speaker NP Ellipsis in Mandarin

while Objects are expressed by the more marked form: nouns. This is consistent with Du Bois' (1987) findings that new participants, which require more identifying information, are introduced first in Object position, while Subjects tend to be established participants, who can thus be referred to with less explicit forms.

Discussion

Native speakers of Chinese use ellipsis of NPs at a much higher frequency than English-speaking learners do. Ellipsis of Subjects accounts for the major part of the difference, with the means for realised and ellipsed Subject arguments being roughly equal for native speakers (12.22, sd 5.1, and 11.33 sd 1.5 respectively, see Table 1) but varying greatly for learners (12.286 sd 3.6, and 3.143 sd 2.7). The Korean-English text was closer to native Chinese texts than learner Chinese, having similar (though not so similar as for native Chinese) frequencies of realised and ellipsed Subject NPs (88 vs 64). The contrast between the Korean speaker of English and the English-speaking learners of Chinese suggests that high frequencies of NP ellipsis are more likely to be the result of L1 transfer than a consequence of 'limited linguistic resources' and the pressures of production in a Second Language.

As stated in the introduction, L1 transfer may influence L2 acquisition in various ways. A typological contrast between L1 and L2 may result in a longer phase of error production in the areas of syntax upon which the typology is based. In Government and Binding Theory definite (optional) ellipsis is known as 'pro-drop' and is considered to be one of a number of parameters which are 'set' in L1 acquisition (Chomsky, 1981). This constitutes a typological distinction between 'pro-drop' and non-'pro-drop' languages. Nonetheless, the constraints on, and functions of ellipsis generally are clearly complex, with some overlap of form and function in English and Mandarin.

Similarly, definite ellipsis cannot be characterised as a 'zero category' in English, as zero pronouns can be employed in English, albeit to a much more limited extent than in Mandarin. English-speaking learners of Mandarin may quickly master the use of ellipsis in familiar (L1) contexts, yet still be slow to employ this strategy in unfamiliar (L2) contexts. That is to say, the difficulty is not one of the acquisition of a form or mode of reference so much as a difficulty of recognising the discourse contexts in which it is appropriate to employ an already familiar strategy. While the various referential forms are easily acquired, being less complex than English pronouns, the overall pattern of ellipsis used by native speakers may be more elusive. The corollary of an underuse of a familiar strategy (zero anaphora), in new contexts, is the over-use, from an L2 viewpoint, of an alternative strategy (lexical pronouns) in a context where it would not be the preferred strategy of a native speaker.

It will be recalled that Eckman's Markedness Differential Hypothesis predicts that it is areas that are **more marked** that will be more difficult to acquire, but Larsen-Freeman and Long (1991) point out that elementary

learners have been known to transfer more marked forms into L2. In the case of English-speakers learning Mandarin, the task is to learn to **omit** NPs where in English they would employ a pronoun. In terms of formal markedness, this is clearly replacing a more marked form, with a less marked one. Clearly, this shift also presents problems for the learners. We have seen that even second year students of Mandarin persist in the use of lexical pronouns, despite the fact that they are **more marked**, morphologically and phonologically speaking, than a zero pronoun.

For Eckman, markedness means more morphologically, phonologically or syntactically complex, but also less frequent (statistically). Use of zero pronouns must necessarily be less marked, in any formal sense, than the use of realised pronouns but from the viewpoint of the English speaker, zero pronouns are more marked, in L1, in terms of frequency, ie less common, than lexical pronouns. It appears that this statistical markedness within the L1 works as a disincentive to production of this form, even though it is morphologically less complex and more frequent from the viewpoint of the target language, Mandarin. Where L1 transfer results merely in over-use of a valid L2 form, rather than in grammatical error, it may persist in the language of advanced learners, even where it constitutes the use of a more marked form in a less marked context.

Complement structures, involving the copula *shì* are slightly more common in learner texts. This may be because they resemble English Subject predicate structures with the verb *be*. In these contexts native speakers tend to use simple juxtaposition of an NP and a stative verb.

It might have been expected that Object ellipsis, being ungrammatical in English, would have proved to be a more significant area of difference between the two groups. However, it is clear that it is a relatively low frequency phenomenon even in the texts of native speakers. This being the case, a proper evaluation of native and learner differences with regard to Object ellipsis, requires larger samples of text. It should also be noted that the overall proportion of Object to Subject NPs is not particularly low in either native or learner texts; roughly 40:60 in both cases. However the vast majority of Objects are expressed by full NPs. Thus the low frequency of ellipsed Objects in native texts is most likely a consequence of the low frequency of Objects which are topical (readily available to the listener/reader), and not a consequence of grammatical constraints on object ellipsis. The non-topical status of Objects is purportedly a discourse universal (Du Bois, 1987), as is the use of more marked forms for less accessible topics (Givon 1984). If so, this does not represent a learning task for acquirers of another language, and this accounts for the similarity between native and non-native frequencies in Object ellipsis.

The fact that learners do not ellipse Subject arguments with the same frequency as natives is most likely to be explained in terms of the syntactic contexts in which each group exercises ellipsis. It should be remembered that ellipsed NPs in this study do not include instances of obligatory ellipsis

A Comparison of Learner and Native Speaker NP Ellipsis in Mandarin

(equi), and in fact there were no instances of learners attempting to use lexical NPs where they are unacceptable in Chinese.

Reading of the texts suggest that this is because such structures in Chinese are paralleled by infinitive or complement structures in English where NPs are also obligatorily omitted. Thus the absence of any special marking on non-finite verbs in Chinese does not appear to constitute a problem for acquisition of ellipsis in sub-ordinate and complement structures by English speakers. NP ellipsis in this context is found in many languages and this may represent an area where Universal Grammar operates to prevent excessive generalisation of language specific anaphoric differences.

There was one instance of ungrammatical omission. This involved ellipsis of the Head of a nominal modifier in a Complement. Addition of the nominal particle DE would have rendered the sentence correct, and the error was probably a mis-handling of this construction. This suggests that learners may be aware of the acceptability of ellipsis, but unpractised in the details of performing it. That is to say, though ellipsis appears to be simply a case of leaving something out, learners may be unclear of just what can be left out where. The precise structural positions in which natives do ellipse and learners do not remains to be analysed. A likely source appears to be in a greater use of serial verb constructions for coordination and subordination in native texts.

Conclusion

The statistical analysis reported above indicates that there are significant differences in the use of Anaphora by native Chinese and second language learners of Chinese, and that these are most likely the result of L1 transfer. Two years of instruction and in some cases substantially more exposure, have not produced native-like patterns of ellipsis in learners. Nonetheless, since obligatory ellipsis does appear to have been 'mastered' even without instruction, the non-native patterns of learners produce noticeable 'stylistic' differences, but do not, by and large, produce grammatical errors. One possible exception is the mastery of the DE phrase with deleted Head. The fact that failure to ellipse does not result in ungrammaticality may explain why subjects with a reasonable level of 'proficiency' persist with non-native patterns.

The difference in levels of ellipsis was found to be expressed most clearly in regard to Subject NPs, since it is Subjects which are most readily accessible and therefore eligible for ellipsis.

Just which syntactic or discourse structures contribute to this difference is the next stage of this investigation. Preliminary analysis of texts suggests that ellipsis in Chinese functions as a coordinating and subordinating device, expressing relationships between predications and not just as a referent tracking device. (See Charters, 1996b for a detailed analysis of native-speaker ellipsis, and Charters 1996a for results of a comparison of the distribution of ellipsis in native and learner texts.) This

could account for Williams' findings that the English of Chinese-speakers exhibits a greater frequency of anaphor in **non-coordinating** structures than native English. It is also consistent with Li and Thompson's (1979) observation that the use of pronominals in native Chinese texts is inversely related to the extent to which clauses are conjoinable.

In this case, ellipsis may be best described in terms of a four-way distinction: **obligatory absence** (zero anaphora) of an argument, indicating co-reference with the Subject of the Matrix clause (equi); **contrastive ellipsis** where absence of an NP does not affect grammaticality, but contrasts referentially with a realised NP (see Huang 1984) and **optional ellipsis** where the presence or absence of an NP affects neither the grammaticality of the structure nor its interpretation. The last includes **conjunctive ellipsis** where presence or absence of an NP does not affect grammaticality, or referentiality but does affect the interpretation of **relationships between predications** e.g. purpose, causality, co-ordination and sub-ordination.

Obviously grammaticality and clear communication should be the main aims of beginning students, but it is to be hoped that for advanced learners, and the teachers who work with them, a close approach to the stylistic conventions of the target language and a native-like command will be equally important goals. One step towards these goals is the production and critical appraisal of texts longer than a single sentence, and the role of ellipsis within them.

References

- Aoun, J. 1985. *A Grammar of Anaphora*. Cambridge, MA: MIT Press.
- Battistella, E. 1985. 'On the distribution of PRO in Chinese', *Natural Language and Linguistic Theory* 3:317-340.
- Bresnan, J. (ed) 1982. *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press
- Charters, A.H. 1996a. *Cut "it" out: the avoidance of ellipsis in written texts by learners of Mandarin*. Paper presented at the Pacific Second Language Research Forum, Wellington, Feb. 1996.
- Charters, A.H. 1996b. *A silence more eloquent: NP ellipsis in Mandarin discourse*. MA thesis (The Australian National University).
- Chen Ping. 1984. *A Discourse Analysis of Third Person Zero Anaphora in Chinese*. Bloomington Indiana: Indian University Linguistics Club
- Chomsky, N. 1980 'On Binding', *Linguistic Inquiry* 11:1-46
- Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, N. 1982, *Some Concepts and Consequences of the theory of Government and Binding*. Cambridge, Mass.; MIT Press.
- Chung, Sandra. 1989. 'On the notion of "null anaphor" in Chamorro.' In Jaeggli and Safir (eds), *The Null Subject Parameter*. Dordrecht: Kluwer. 143-184.
- Du Bois, J. 1987. 'The Discourse Basis of Ergativity', *Language* 63 (4): 805-855.
- Eckman, Fred R. 1977. 'Markedness and the Contrastive Analysis Hypothesis', *Language Learning* 27:315-330.
- Eckman, Fred R. 1985. 'Some theoretical and pedagogical implications of the

A Comparison of Learner and Native Speaker NP Ellipsis in Mandarin

- markedness differential hypothesis', *Studies in Second Language Acquisition* 7:289-307.
- Fakhri, A. 1984. 'The use of communicative strategies in narrative discourse: a case study of a learner of Moroccan Arabic', *Language Learning*, 34 (3): 15-37.
- Farmer A.K. and R.M. Harnish. 1987. 'Communicative reference with pronouns.' In Verschueren & Bertuccelli-Papa (eds) *The Pragmatic Perspective*. Amsterdam: John Benjamins. 547-65.
- Fillmore, Charles J. 1986. 'Pragmatically controlled zero anaphora.' In Nikiforidou et al. (eds) *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*. Berkeley: BLS. 95-107.
- Givon, T. 1983. 'Introduction.' In T. Givon (ed.) *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Vol. 3 of *Typological Studies in Language*. Amsterdam: Benjamins, pp. 5-39.
- Givon, T. 1984. 'Universals of Discourse Structure and Second Language Acquisition.' In W. Rutherford (ed.) *Language Universals and Second Language Acquisition*. Amsterdam: John Benjamins, pp. 109-33.
- Gundel, J. and E. Tarone. 1983. 'Language Transfer and the Acquisition of pronominal anaphora.' In S. Gass and L. Selinker (eds) *Language Transfer and Language Learning*. Rowley, MA: Newbury House, pp. 281-96.
- Haegeman, L. 1991. *Introduction to Government and Binding Theory*. Oxford/Cambridge Mass: Basil Blackwell.
- Huang, C.T. James. 1982. *Logical relations in Chinese and the theory of grammar*. Ph.D. dissertation. MIT.
- 1984. 'On the distribution and Reference of Empty Pronouns', *Linguistic Inquiry* 15(4) 531-44
- 1989. 'Pro-drop in Chinese: a Generalised Control Theory.' In Jaeggli and Safir (eds) *The Null Subject Parameter*. Dordrecht/Boston/ London: Kluwer Academic Publishers.
- 1993. 'Reconstruction and the structure of VP: some theoretical consequences', *Linguistic Inquiry* 24(1) 103-138.
- Klein, W. 1986. *Second Language Acquisition*. Cambridge: Cambridge University Press.
- Huang Yan. 1992. 'Against Chomsky's typology of Empty Categories', *Journal of Pragmatics* 17: 1-29.
- Huang Yan. 1994. *The syntax and pragmatics of anaphora*. Cambridge: Cambridge University Press.
- Kaplan, Ronald, M. and Joan Bresnan. 1982. 'Lexical-Functional Grammar: A Formal System for Grammatical Representation.' In Bresnan (ed) *The Mental Representation of Grammatical Relations*. pp. 174-281.
- La Polla, R. 1993. 'Arguments Against "Subject" and "Object" as Viable Concepts in Chinese', *The Bulletin of the Institute of History and Philology, Academia Sinica*. Vol LXIII Part IV. Taiwan: Academia Sinica.
- 1994. 'Pragmatic relations and word order in Chinese.' In M. Noonan and P. Downing (eds) *Word Order in Discourse* (Typological Studies in Language 30. Amsterdam/Philadelphia: John Benjamins. pp 299-331.
- Larsen-Freeman, D. and M.H. Long. 1991. *An introduction to Second Language Acquisition Research*. London and New York: Longman.
- Levinson, S.C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- 1987a. 'Minimization and conversational inference.' In Verschueren and Bertuccelli-Papi (eds) *The Pragmatic Perspective*. Amsterdam: John Benjamins. pp. 61-129.

- 1987b. 'Pragmatics and the grammar of anaphora: a partial pragmatic reduction of Binding and Control phenomena', *Journal of Linguistics* 23, 379-434.
- 1991. 'Pragmatic reduction of the Binding Conditions revisited', *Journal of Linguistics* 27. 107-61.
- Li, Charles and Sandra Thompson. 1979. 'Third-Person Pronouns and Zero-Anaphora in Chinese Discourse.' In Talmy Givon (ed) *Syntax and Semantics* Vol 12. Discourse and Syntax. New York and London: Academic Press
- Lust, Barbara. 1986. (ed) *Studies in the Acquisition of Anaphora Vol I, defining the constraints*. Dordrecht: D. Reidel.
- McLuhan, M. 1964. *Understanding media: the extensions of man*. New York: McGraw Hill.
- Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London/New York: Longman.
- Reinhart, Tanya. 1976. *The Syntactic Domain of Anaphora* (PhD dissertation, MIT, Cambridge, Mass.)
- 1983. *Anaphora and Semantic Interpretation*. London and Canberra: Croom Helm.
- Rizzi, L. 1986. 'Null Objects in Italian and the theory of pro', *Linguistic Inquiry*, 17, 501-58.
- Shopen T. 1972. *A Generative Theory of Ellipsis: A Consideration of the Linguistic Use of Silence*. Reproduced by the Indiana University Linguistics Club.
- 1973 'Ellipsis as Grammatical Indeterminacy', *Foundations of Language* 10:65-77.
- White, L. 1985. 'The "pro-drop" parameter in adult second language acquisition', *Language Learning* 35: 47-61.
- Williams, J. 1989. 'Pronoun copies, pronominal and zero anaphora.' In S. Gass et al (eds) *Variation in Second Language Acquisition: Discourse and Pragmatics*. Avon: Multilingual Matters. pp. 153-189.