A corpus-based study of Mandarin Chinese referring expressions in oral narratives of preschool children

Yifang Yuan
Simon Fraser University
yifang_yuan_2@sfu.ca

This study utilizes a corpus-based approach to investigate the use of referring expressions, such as definite and indefinite noun phrases and pronouns, in oral narratives produced by monolingual Mandarin Chinese-speaking preschool children. The data for this study were collected from the spoken narratives of five 4-year-old children selected from the Zhou Narratives corpus (Li & Zhou, 2011). Adopting the cognitive approach of Gundel et al. (1993), this study analyzes the relationship between the forms of referring expressions and their corresponding cognitive statuses and discourse functions. The results indicate that the correlations between the referential forms produced by 4-year-old Chinese monolingual children and their cognitive statuses and discourse functions align with the predictions of the Givenness Hierarchy and the patterns observed in earlier research conducted by Gundel et al. (1993) with Chinese adults. However, this study also reveals notable differences in the preferred referential forms used by Chinese monolingual children and adults in relation to specific cognitive statuses and discourse functions, suggesting that the development of referential appropriateness in narrative production follows a gradual trajectory in children.
Keywords: Mandarin Chinese; referring expressions; cognitive status; discourse function

1   Introduction

Referring expressions (REs), such as definite and indefinite noun phrases, demonstrative noun phrases, and overt and zero pronouns, play a critical role in facilitating effective communication by providing processing signals for addressees to identify the intended referent of a speaker.[1] These forms are essential in narratives, fulfilling discourse functions such as introducing, re-introducing, and maintaining referents (Chen & Lei, 2013). It has long been

---

[1] Abbreviations used in this paper are ø 'null pronoun', ART 'article', ASP 'aspect', CL 'classifier', COP 'copula', DEM 'demonstrative', DEM1 'proximal demonstrative', DEM3 'distal demonstrative', EXT 'existential', SG 'singular', N 'noun phrase', NEG 'negation', NP 'noun phrase', OBJ 'object', PL 'plural', POSS 'possessive', PRO 'pronoun', PROG 'progressive aspect', Q 'question marker', QUANT 'quantifier', RE 'referring expression', and SUB 'subject'.

observed that human languages often use different expressions to refer to the same object, and a single expression can also be used to refer to different objects. For example, in English, a particular boy can be referred to as *a boy*, *the boy*, *this boy*, *that boy*, *this*, *that*, or *he*. This phenomenon raises one of the fundamental issues in the field of reference studies, that is, how speakers choose the appropriate form to refer to a particular object and how addressees are able to correctly identify the intended referent despite the potential ambiguity of the expression.

The inventory of referring expressions varies from language to language. In Chinese, common forms of referring expressions include the proximal demonstrative *zhe*, distal demonstrative *na*, indefinite article *yi* 'one', bare nouns, overt pronouns such as *ta* '3SG', and zero pronouns. Unlike English, Chinese commonly uses bare nouns, as the language lacks distinctive forms for articles. Chinese lacks a definite article, and the indefinite article *yi* 'one' is optional. Moreover, as a pro-drop language, Chinese allows for the omission of subjects or objects in certain contexts, resulting in the frequent use of zero pronouns in discourse. Furthermore, the relationship between the forms of referring expressions and their corresponding cognitive statuses and discourse functions varies across languages. This variability can be seen in the comparison of the following Chinese data from the Golden Fish Corpus (Fuller & Gundel, 1987) with its English translation.[2]

(1) Zhege   shihou   you   yi-ge   haizi   qu   dushu.
     DEM1   *time*   *EXT*   *one-CL*   *child*   *go*   *study*
     ø   huidao   jia-li   yihou,   ø   kandao   zhuozi-shang
     (RE) *return* *home-in*   *after*   *(RE)* *see*   *table-above*
     fangzhe   liang-ge   tongbi   he   yi-ge   pingzi.
     place   *two-CL*   *coin*   *and*   *one-CL*   *bottle*
     'At this time, there was a child going to school. When (he) came home, (he) saw on (the) table two coins and a bottle.'

The example in (1) illustrates that both Chinese and English use a proximal demonstrative determiner to refer to the stage topic 'this time', and an indefinite article to introduce a discourse-new entity, namely, the main character of the story. However, Chinese and English differ in the forms of referring expressions used to maintain reference to story characters. When referring to the boy who has been introduced in the preceding utterance, Chinese uses a zero pronoun while English uses an overt pronoun. Furthermore, both languages use numerals to

---

[2] The Golden Fish Corpus (Fuller & Gundel, 1987) is a collection of forty-five narratives that were elicited from twenty-five native speakers of English, Chinese, Japanese, Korean, Arabic, Spanish, and Farsi. The participants were shown a twenty-minute film titled "The Golden Fish" (Séchan, 1959), which features a storyline accompanied solely by a musical soundtrack. After watching the film, the participants were instructed to narrate the story in English to a native English speaker and then retell it in their native language to another speaker of that language.

introduce plural objects ('two coins'), and indefinite articles to introduce singular objects ('a bottle').

Previous research on Chinese referring expressions has mainly focused on aspects such as form and distribution (van Deemter et al., 2017), discourse function (Yang et al., 1999), information structure (Nie, 2020), cognitive status (Gundel et al., 1993; Shi, 1998), and interpretation of certain referential forms (Kuo, 2008). These studies primarily relied on data from the discourse of Chinese adults. Although some previous studies have investigated the development of referring expressions in the discourse of Chinese-English bilingual children (Chen & Lei, 2013; Chen & Pan, 2009), there has been limited attention devoted to investigating the development of referential appropriateness in monolingual Mandarin Chinese-speaking children, particularly with regards to the use of different forms of referring expressions and their corresponding cognitive statuses and discourse functions in their discourse.

To address this research gap, the present study employs a corpus-based approach to analyze the distribution, discourse function, and cognitive status of referring expressions used in the discourse of monolingual Chinese-speaking children. The data used in this study are extracted from the Zhou Narratives corpus (Li & Zhou, 2011) in the CHILDES database. The Zhou Narratives corpus comprises spoken narratives collected from 200 preschool children. These narratives were elicited using two picture books, namely, *The Very Hungry Caterpillar* (Carle, 1969) and *The Three Robbers* (Ungerer, 1962). The corpus includes both video and text files, accessible for retrieval at https://sla.talkbank.org/TBB/childes/Chinese/Mandarin/ZhouNarratives. The primary objective of this study is to provide a systematic and data-supported analysis of the use of referring expressions in oral narratives produced by Chinese preschool children, and enhance our understanding of the developmental trajectory of referential appropriateness in the communication of young children.

This paper is organized as follows. Section 2 begins by introducing the Givenness Hierarchy and the cognitive approach to referring expressions proposed by Gundel et al. (1993) (§2.1). It also provides an overview of the previous studies on Chinese referring expressions (§2.2). Section 3 describes the data (§3.1) and the criteria used for categorizing and coding Chinese referring expressions in the corpus study (§3.2). Section 4 presents and provides the analysis of the quantitative results obtained from the corpus study. Section 5 concludes the paper.

2   Theoretical background

2.1   The cognitive approach to referring expressions

This section will review Gundel et al.'s (1993) cognitive approach to referring expressions and the Givenness Hierarchy. These concepts will serve as the theoretical framework for explaining the relationship between the form and cognitive status of Chinese referring expressions in this study. The cognitive

approach assumes that different forms of referring expressions conventionally encode different cognitive statuses. Speakers must consider the assumed cognitive status of a referent when selecting an appropriate referring expression, which enables the addressee to identify the intended referent among all the possible options. Based on this assumption, Gundel et al. (1993) propose six implicationally related cognitive statuses to explain the conditions governing the appropriate use of different referring expression forms in natural discourse across five languages–English, Japanese, Mandarin Chinese, Russian, and Spanish. The six cognitive statuses are illustrated in the Givenness Hierarchy depicted in Table 1 (Gundel et al., 1993, p. 284).

Table 1. *The Givenness Hierarchy with relevant RE forms in English and Chinese*

	In focus	Activated	Familiar	Uniquely identifiable	Referential	Type identifiable
English	*it*	*HE* *that* *this* *this* N	*that* N	*the* N	indefinite *this* N	*a* N
Chinese	ø *ta* (3SG)	*TA* *zhe* 'this' *nei* 'that' *zhe* N		*nei* N		*yi* N 'a N' ø N

Note. The relevant forms in the table represent members of a whole class. *It* in the 'English' 'in focus' cell represents all phonetically unstressed personal pronouns. *HE* in the 'English' 'activated' cell represents all phonetically stressed personal pronouns (Gundel et al., 1993, p. 284).

The Givenness Hierarchy provided in Table 1 is illustrated with the relevant forms of referring expressions in both English and Mandarin Chinese. This table simplifies the representation by including only one form to represent members of a whole class. In English, for example, the demonstrative determiners *that* and *this* also represent the forms *those* and *these*, respectively. This hierarchical framework correlates the form of referring expressions with six distinct cognitive statuses of the reference: *type identifiable*, *referential*, *uniquely identifiable*, *familiar*, *activated*, and *in focus*. The specific meanings of these statuses will be discussed in detail in the remaining portion of this subsection. These six cognitive statuses represent the assumptions made by the speaker regarding the addressee's knowledge and attention state. Each status is a necessary and sufficient condition for the appropriate use of specific referring expression form(s). To illustrate, in Chinese, the utilization of a zero pronoun (ø) or an unstressed pronoun requires that the referent is in focus, indicating that it is at the center of attention in the current discourse.

It should be noted that not all six cognitive statuses are necessary for all languages. As indicated in Table 1, English has a form (indefinite *this* N) for

which the status "referential" is both necessary and sufficient. However, Chinese lacks forms that specifically require the referent to be referential, but not necessarily uniquely identifiable. The cognitive statuses in the Givenness Hierarchy are implicationally related in the sense that they are ordered on a restrictiveness continuum with regard to the set of possible referring expression forms they are associated with. Each status in the model entails all lower statuses but not vice versa, which means that a particular form can be substituted by any form that requires a lower status (to the right) in the hierarchy. For example, in English, a referent that is uniquely identifiable can appropriately be referred to using either *a* N, indefinite *this* N, and *the* N.

In the Givenness Hierarchy, the status of "type identifiable" implies that the addressee is able to access a representation of the type of object denoted by the referring expression. In Chinese, this status is sufficient for the use of the indefinite article *yi* 'one' and bare nouns. The status of "referential" indicates that the addressee can either retrieve an existing representation or construct a new one of the intended referent after processing the complete sentence. In Chinese, this status is not correlated with any particular expressions. However, in English, it is both necessary and sufficient for the use of the indefinite demonstrative determiner *this.* The status of "uniquely identifiable" denotes when the addressee can identify the intended referent based solely on the referring expression itself. This status is both necessary and sufficient for the appropriate use of the distal demonstrative determiner *na* 'that' in Chinese and the definite article *the* in English. The crucial difference between the statuses of "referential" and "uniquely identifiable" lies in the fact that the former requires the addressee to construct a representation of the referent based on the nominal along with the content of the rest of the sentence, while the latter relies solely on the nominal itself. The status of "familiar" indicates that the addressee already possesses a representation of the referring expression in either long-term or short-term memory, enabling them to uniquely identify the intended referent. Chinese does not have any referential forms specifically associated with the status of "familiar". However, in English, this status is both necessary and sufficient for the appropriate use of the distal demonstrative determiner *that.*

The status of "activated" implies that the referent exists in the addressee's current short-term memory, either through retrieval from long-term memory or arising from the immediate context. This status is both necessary and sufficient for appropriate use of stressed personal pronouns, demonstrative pronouns, and the proximal demonstrative determiner in both Chinese and English. Finally, the status of "in focus" is the most restrictive status in the hierarchy. It indicates that the referent is at the current center of attention. This status is both necessary and sufficient for the appropriate use of zero pronouns and unstressed pronouns in both English and Chinese. In natural discourse, referents in focus often coincide with the topic of the preceding utterance, and consequently, they can be partially anticipated through the syntactic structure of the referring expression. Nevertheless, the actual cognitive statuses of referents are ultimately determined by pragmatic factors. To demonstrate the correlation between the various forms

of referring expressions and the conditions for their appropriate use and interpretation in Chinese, consider the following examples.

(2) **Yi-tiao-gou/gou**   chao-de   wo   shui   bu   zhao.
    one-CL-dog/dog   *disturb-PRT*   *1SG*   *sleep*   *NEG*   *PRT*
    'A noisy dog keeps me awake.'   [TYPE IDENTIFIABLE]

Unlike English, Chinese does not differentiate between all six cognitive statuses. It lacks distinctive forms of referring expressions for which the statuses of "referential" and "familiar" are both necessary and sufficient. In Chinese, both bare nouns and indefinite noun phrases can be interpreted as either "type identifiable" or "referential". Moreover, there is no particular referring expression in Chinese that requires the referent to be at most familiar but not activated. For example, in (2), both the definite noun phrase *yi-tiao-gou* 'one-CL-dog' and the bare noun *gou* 'dog' are felicitous in a context where the addressee is assumed to be able to either construct a representation of the dog ('referential') or understand that the speaker is simply asserting the existence of such dog ('type identifiable').

(3) **Na-(tiao)-gou**   chao-de   wo   shui   bu   zhao.
    DEM3-(CL)-dog   *disturb-PRT*   *1SG*   *sleep*   *NEG*   *PRT*
    'The noisy dog keeps me awake.'   [UNIQUELY IDENTIFIABLE]

In Example (3), the use of the distal demonstrative *na* is felicitous as long as the addressee understands that the speaker intends to refer to their neighbor's dog, regardless of whether the addressee has previous knowledge of the dog. In English, the use of the distal demonstrative *that* presupposes that the addressee is at least familiar with the intended referent. Therefore, the Chinese distal demonstrative *na* seems to correspond more closely to the definite article *the* in English rather than the distal demonstrative *that* (Gundel et al., 1993; Li & Thompson, 1981).

(4) **Zhe-(tiao)-gou/ta**   chao-de   wo   shui   bu   zhao.
    DEM-(CL)-dog/3SG   *disturb-PRT*   *1SG*   *sleep*   *NEG*   *PRT*
    'This noisy dog/it keeps me awake.'   [ACTIVATED]

In Example (4), the use of the proximal demonstrative determiner *zhe* 'this' and the stressed pronoun *ta* '3SG' is appropriate only when the referent is represented in the addressee's current short-term memory, either through retrieving from long-term memory or arising from the immediate context.

(5) Wo   linju-de   gou   hen   chao.
     1SG   neighbor-POSS   dog   very   noisy
     Ta/ø   chao-de   wo   shui   bu   zhao.
     3SG/(RE)   disturb-PRT   1SG   sleep   NEG   PRT
     'My neighbor's dog is very noisy. It keeps me awake.'   [IN FOCUS]

In Example (5), the unstressed pronoun *ta* '3SG' and the zero pronoun are felicitous since the referent serves as both the topic and the subject of the preceding sentence. Furthermore, it occupies the current center of attention in the discourse.

2.2   Previous studies on Chinese referring expressions

This section will review some previous studies on Chinese referring expressions that have explored aspects including the distribution, interpretation, discourse function, information structure, and cognitive status of certain referential forms (see e.g. Chen & Lei, 2013; Chen & Pan, 2009; Gundel et al., 1993; Kuo, 2008; Nie, 2020; Shi, 1998; van Deemter et al., 2017; Yang et al., 1999). To investigate the distribution of referring expressions in Chinese natural discourse, van Deemter et al. (2017) focus on the three canonical patterns of reference in Chinese: (1) demonstrative + (classifier) + noun phrase (e.g., *na* (*ge*) *laoren* 'that old person'), (2) bare noun (e.g., *laoren* 'old person'), and (3) indefinite noun phrase (*yi* (*ge*) *laoren* 'an old person'). The findings of their study reveal several noteworthy patterns. First, bare nouns are the most frequently used form, while no instances of demonstratives are found in their dataset. Second, the frequency of preverbal indefinite noun phrases is nearly twice that of their postverbal counterparts. In contrast, the occurrence of preverbal and postverbal bare nouns is approximately equal. This observation suggests that syntactic structure seems to play a role only for the distribution of indefinite noun phrases, but not for bare nouns. Moreover, this finding challenges the traditional views that Chinese preverbal noun phrases take a default definite interpretation, while postverbal noun phrases generally carry an indefinite interpretation. Consequently, indefinite noun phrases and bare nouns are generally restricted to postverbal positions rather than preverbal positions (Chao, 1968/2011).

Referring expressions are generally considered to provide crucial links that integrate successive utterances, thus playing a significant role in promoting discourse coherence. To examine the influence of referring expressions on the comprehension of Chinese discourse, Yang et al. (1999) conduct a series of self-paced reading time studies and compare the processing time required for reduced referential expression (e.g. overt pronouns and zero pronouns) and unreduced expressions (e.g. proper names). The results of their study reveal that sentences containing reduced referring expressions were processed faster than matched sentences with repeated names. This finding suggests that reduced referring expressions in Chinese contribute more to discourse coherence compared to unreduced expressions. Furthermore, their results also indicate that there is no

significant difference in the processing time required for overt pronouns and zero pronouns, which suggests that the two types of reduced referring expressions contribute equally to discourse coherence in Chinese. This finding contradicts the traditional perspective that zero pronouns and overt pronouns perform different roles in pro-drop languages like Chinese, with zero pronouns requiring referents to be more accessible than overt pronouns (Givon, 1983; Ariel, 1991).

In addition to discourse functions, researchers also explore the information structures of Chinese referring expressions. The distribution of Chinese referring expressions is generally constrained by the universal given-before-new order of information structure. Accordingly, preverbal noun phrases usually represent old information and take definite interpretations, while postverbal noun phrases often introduce new information and have indefinite interpretations (Chao, 1968/2011). However, deviations from the definiteness restrictions are not rare in Chinese natural discourse. To examine the non-canonical distribution of Chinese referring expressions in natural discourse, Nie (2020) investigates three Chinese referential choices that do not follow the universal definiteness constriction: bare nouns, indefinite subjects marked by *yi*(-CL) 'one-CL', and existential *you*-sentence with definite objects. The results of Nie's (2020) study reveal that: (i) Chinese bare nouns are typically used to represent thematically unimportant entities, and thus they are less likely to occur in topic chains and exhibit low topicality. Bare nouns can refer to either new or old entities in Chinese discourse. When referring to new entities, they recur significantly less frequently in subsequent discourse compared to the full forms, such as indefinite noun phrases and definite demonstrative noun phrases. When representing old entities, bare nouns tend to exhibit a greater distance from their antecedents in prior discourse compared to the full forms. (ii) Indefinite subjects marked by *yi*(-CL) 'one(-CL)' are licensed by containing relatively old information and the presence of modification, and thus the given-before-new order is partially maintained. Similar to bare nouns, indefinite subjects often refer to thematically unimportant entities and are frequently used as a source of quotation to provide background information. (iii) Definite objects in existential *you*-sentences can refer to either hearer-new or hearer-old entities, as long as their referents do not contain older information than the subjects. Definite objects In *you*-sentences tend to encode background information and refer to entities of low thematic importance that are discontinuous in the subsequent discourse, exhibiting similarities with bare nouns and indefinite subjects. Overall, these three non-canonical structures all serve to signal a low level of topicality and persistence of the referent, and thus they only partially adhere to the universal given-before-new order of information structure.

The previous research on Chinese referring expressions discussed so far has predominantly focused on adult language use. However, there is a growing interest in understanding the development of referring expressions in children's discourse. Chen and Pan (2009) used the picture book *Frog, where are you?* (Mayer, 1969) to investigate the production of English referring expressions in narratives elicited from sixty Chinese children who were learning English as a second language. Subsequently, using the same picture book, Chen and

Lei (2013) conducted a comparative study analyzing referring expressions produced by Chinese-English bilingual children and their monolingual peers in both English and Chinese. Overall, the results of both studies demonstrate that referential appropriateness is developed gradually for both monolingual and bilingual children, and it is influenced by factors such as the discourse function (introduction, re-introduction, or maintenance of referents) and the character type of the referent (main or secondary story character).

In this section, we reviewed the literature on Chinese referring expression studies from the perspectives of their distribution, discourse function, information structure, and development of production. The following section will introduce the methodology, data material, and coding and analysis procedures utilized in the present study.

3   The present study

3.1   Methods

Utilizing a corpus-based approach, this study investigates the production of referring expressions in the narratives of five 4-year-old Chinese monolingual children. The spoken narratives analyzed in this study are extracted from the Zhou Narratives corpus (Li & Zhou, 2011), including both video files and corresponding transcribed text files, from the CHILDES database. The Zhou Narratives corpus comprises spoken narratives collected from 200 Chinese-speaking preschool children in 2008, elicited using two picture books: *The Very Hungry Caterpillar* (Carle, 1969) and *The Three Robbers* (Ungerer, 1962). Specifically, the data analyzed in this study are selected from the subset of the corpus that was elicited using the latter book. Considering the existing body of research that has explored the production of referring expressions in children starting from the age of 5 years, the present study restricts its scope to include 4-year-old children, with the aim to investigate the use of referring expressions in younger age groups. The elicitation in the Zhou Narratives corpus was conducted in a quiet room at the participants' school, with each child interviewed individually. The participants were instructed to familiarize themselves with the entire book and then retell the story to a toy that had not heard the story before. Minimal instructions, such as prompting with *What happened next?*, were provided only when the children seemed to have trouble producing narration at any point.

3.2   Coding and analysis procedures

The referring expressions occurring in children's speech were identified and coded in terms of their grammatical forms, discourse functions, definiteness, and syntactic position. Table 2 presents the format of the coding scheme used in this study along with an example from the Zhou Narrative corpus. The "code" row in Table 2 encodes the filename of the recording, while the "utterance" row encodes

the transcript of the utterance containing the target referring expression highlighted in bold. The "type" row represents the general form of the referring expression, such as *yi*-N in Example (6), which stands for noun phrases with the indefinite determiner *yi* 'one'. However, in natural data, the form of referring expressions can be further classified according to their actual grammatical form. For example, the form of *yi*-N includes both noun phrases with the indefinite determiner *yi* and noun phrases with *yi* followed by a classifier (CL).

Table 2. *Coding scheme for referring expressions*

Code	4yf01			
Utterance	(6) **Yi-ge**	**da**	**ma**	guolai-le
	one-CL	*big*	*horse*	*come-ASP*
	'Here comes a big horse.'			
Type	*yi*-N			
Form	*yi*-CL-N			
Cognitive status	Type identifiable			
Discourse function	Introduction			
Position	Subject			
Gesture	N/A			

The actual grammatical form of the referring expression is stored in the "form" row, including categories such as *yi*-CL-N. The "cognitive status" row indicates the interpretation of the cognitive status of the referring expression, following Gundel et al.'s (1993) classification criteria. The "discourse function" row captures the specific role of referring expressions in relation to introducing, re-introducing, or maintaining reference to story characters. Following Serratrice (2007) and Chen and Lei (2013), this study codes the discourse function of referring expressions that are used to introduce a story character to the discourse as "introduction". All referring expressions that refer to entities not mentioned in the immediately preceding clause and/or express a subject argument whose immediate antecedent occurs in object position are coded for the discourse function of "re-introduction". The discourse function of all other subsequent mentions of story characters is coded as "maintenance". The "position" row records the syntactic information of the referring expression (e.g., subject and object). Finally, the "gesture" row notes any observed gestures, such as pointing to images in the picture book.

(7) a. **yi-ge-ta**   [introduction]
         one-CL-PRO
     b. **zhe-ge-ta**   [maintenance]
         DEM-CL-PRO
     c. ta-de   mama   bao-zhe   **ta**   [maintenance]
         PRO-POSS   *mother*   *hold-PROG*   *PRO*

d. **ta**   zai   shuijiao   ne   [re-introduction]
    PRO *PROG* *sleep*   *PRT*
e. **ta**   ba   ta-de   haizi   touzou   le   [re-introduction]
    PRO *ba*   *PRO-POSS*   *child* *steal*   *PRT*
    Lit: 'One she, this she, her mother was holding her. She was sleeping.
    He stole her child.'

In (7), an excerpt from the recording of 4yf02 is provided to exemplify how to code the discourse function of referring expressions in this study. Note that the highlighted referring expressions in (7a) to (7d) refer to a girl in the story, while the pronoun in (7e) refers to a robber who steals the girl from her mother. The discourse function of the targeted referring expression in (7a) is coded as "introduction" since its referent is introduced for the first time in the discourse. Both (7b) and (7c) are coded as "maintenance" because they refer to the same referent as (7a). In (7d), the discourse function is coded as "re-introduction" as it refers to a subject argument in (7d) whose immediate antecedent in (7c) appears in the object position. Finally, (7e) is coded as "re-introduction" because its referent has been mentioned in previous discourse, but not in the immediately preceding clause.

To gain a comprehensive understanding of the distribution of referring expressions in the speech of Chinese monolingual children, we examine each type of referential form found in our data in terms of the cognitive status, discourse function, and sentence structure. The next section will present the quantitative results of the corpus study and provide an analysis of the findings.

4   Results and discussion

In Table 3, we present the frequency of occurrence of different types of Chinese referring expressions in our dataset, ranked in descending order based on their total occurrences (raw frequencies). The column labeled "Freq." displays the frequency count, accompanied by the corresponding percentages in parentheses. For comparison purposes, we have reproduced the results of the distribution of Chinese referring expressions from Gundel et al.'s (1993) study in the table.

Table 3. *Distribution of referring expressions from the recordings of 4yf01–05 in the Zhou Narrative corpus*

Type	Rank	Freq. (%)	Rank	Freq. (%) (Gundel et al. 1993)
N	1	157 (37.0%)	1	104 (43.3%)
PRO	2	79 (18.6%)	2	40 (16.7%)
ø	3	53 (12.5%)	4	26 (10.8%)
zhe	4	43 (10.1%)	7	2 (0.8%)
QUANT-N	5	32 (7.5%)		N/A
zhe-N	6	25 (5.9%)	3	39 (16.3%)
yi-N	7	22 (5.2%)	5	19 (7.9%)
na-N	8	12 (2.8%)	6	10 (4.2%)
na	9	1 (0.2%)	8	0
Total		424 (100%)		240 (100%)

As shown in Table 3, the referring expressions in our corpus data totals to 424 tokens. Among the nine different forms of referring expressions, bare nouns (N) have the highest raw frequency, occurring 157 times and accounting for 37.0% of the total number of referring expressions in our data. The second most frequently used type of referring expression is overt pronoun (PRO), occurring 79 times (18.6%) in total, followed by zero pronoun (ø) which occurs 53 times (12.5%), the proximal demonstrative pronoun *zhe* (43 times, 10.1%), noun phrases modified by quantifiers (e.g. *xuduo* 'many') or numerals (e.g. *san* 'three') (32 times, 7.5%), noun phrases with the proximal demonstrative determiner *zhe* (25 times, 5.9%), noun phrases with the indefinite article *yi* 'one' (22 times, 5.2%), and noun phrases with the distal demonstrative determiner *na* (12 times, 2.8%). The raw frequency of the distal demonstrative pronoun *na* is significantly lower than other forms, occurring only once in our dataset.

The distribution of referring expressions in the present study generally follows the pattern observed in Gundel et al.'s (1993) study. Bare nouns and overt pronouns are the two most frequently used forms in both studies. Note that noun phrases modified by quantifiers or numerals are not included in Gundel et al.'s (1993) study. Noun phrases with the indefinite article *yi* and distal demonstrative determiner *na*, and the distal demonstrative pronoun *na* all rank low in both studies. The frequencies of *yi*-N (5.2% vs. 7.9%) and *na*-N (2.8% vs. 4.2%) are both slightly lower in this study than in the previous study. Only one distal demonstrative pronoun *na* occurs in our dataset, and no occurrences of *na* were found in the previous study. The most noticeable difference between the distribution of referring expressions in this study and that of the previous study lies in the proximal demonstrative *zhe*. The proximal demonstrative pronoun *zhe* occurs 43 times and represents 10.1% of the total number of referring expressions in our data, whereas only two instances (0.8%) of *zhe* were found in the previous study. In contrast, a opposite pattern is found for noun phrase with the proximal demonstrative determiner *zhe*, which account for 5.9% (25 times) in our study but 16.3% (39 times) in the previous study.

The results of the distribution of referring expressions according to their cognitive statuses are presented in Table 4. For ease of comparison, we have reproduced the results from Gundel et al.'s (1993) study in Table 5.

Table 4. *Distribution of referring expressions in the recordings of 4yf01–05 from the Zhou Narrative corpus according to cognitive statuses*

	In focus	Activated	Familiar	Uniquely identifiable	Referential	Type identifiable	Total
ø	40	13					53
PRO	46	31			2		79
zhe	27	16					43
na		1					1
zhe N	12	13					25
na N	4	8					12
yi N	2	1				19	22
N	18	49		12		78	157
QUANT-N	3	9				20	32
Total	152	141		14		117	424

Table 5. *Distribution of referring expressions according to cognitive statuses from Gundel et al. (1993)*

	In focus	Activated	Familiar	Uniquely identifiable	Referential	Type identifiable	Total
ø	25	1					26
PRO	40						40
zhe		2					2
na							
zhe N	12	26	1				39
na N	1	7	2				10
yi N					17	2	19
N	12	17	14	49	2	10	104
Total	90	53	17	49	19	12	**240**

Overall, the distribution of referring expressions according to their cognitive statuses in the present study follows the results of Gundel et al.'s (1993) study. Almost all forms of referring expressions occurring in our dataset meet their corresponding necessary conditions predicted by the Givenness Hierarchy. Most of the zero pronouns (40 occurrences, 75.5%) are used for referents that are in focus, as exemplified in (8). Almost all the overt pronouns found in our data refer to entities that are at least activated, as in (9) and (10). The demonstrative determiners *zhe* and *na* never code referents that are familiar, uniquely identifiable, referential or type identifiable. The distal demonstrative pronoun *na* rarely occurs in our dataset. The only occurrence of *na* found in our data is

provided in (11). Finally, all referents of indefinite noun phrases, bare nouns, and noun phrases with quantifiers or numerals are at least type identifiable.

(8) Xiao-guaiwu-men   ye   dao   loushang   [IN FOCUS]
small-monster-PL   also   towards   upstairs
qu   shai-le   yifu. Ranhou   ø   ba
go   hang-ASP   cloth   afterwards   (RE)   BA
na-ge-xiao-nühai   fangzai   ø   jia   menkou
DEM-CL-small-girl   put   (RE)   home   doorway
'(The) little monsters also went upstairs and hanged (the) clothes. Afterwards, (they) put the little girl in front of (her) house.'

(9) Zhe-xie   shi   hao   ren.   [IN FOCUS]
DEM-PL   COP   good   people
Ta-men   ba   yifu, kuzi   nongde   lande   le.
3SG-PL   BA   cloth   pant   make   blue   PRT
'These are good people. They got (their) clothes and pants blue.'

(10) **Ta-men**   zai   tou   dongxi   ne   [ACTIVATED]
3SG-PL   PROG   steal   thing   PRT
'They are stealing something.'

(11) Zhe-li   sange-qiangdao   dao   [ACTIVATED]
DEM-in   three-robber   toward
mofashi   **na-li**   qu   le.
magician   DEM-in   go   PRT
'Here, (the) three robbers went to (the) magician.'

However, the results of the present study also show some patterns that do not follow the results of the previous study. For example, 24.5% (13 occurrences) of the zero pronouns found in our data are used for activated referents, as in (12), compared to the 3.8% (1 occurrence) of zero pronouns in Gundel et al.'s (1993) study.

(12) Zheli   ø   you   bao-qi-le   xiao-nühai.   [ACTIVATED]
DEM   (RE)   again   hold-up-ASP   small-girl
'Here, (he) held (the) little girl again.'

There are 31 occurrences (39.2%) of overt pronouns in our data that are used for activated referents, as exemplified in (13), and 2 occurrences are referential, as shown in (14), whereas all overt pronouns in the previous study are in focus.

(13) Xiao-fangchui   hai   you   ju-ren   [ACTIVATED]
     small-spindle   also   EXT   huge-people
     yao   qiao   **ta-men**   yixia.
     want   hit   3SG-PL   once
     '(The) small spindle and (the) giant want to hit them once.'

(14) **Ta**   shi   ta-men-de   laoshi.   [REFERENTIAL]
     3SG   is   3SG-PL-POSS   teacher
     'He is their teacher.'

The pronoun in (13) is coded for the cognitive status of activated because its referent has been mentioned in the previous discourse, but not in the immediately preceding utterance. The pronoun in (14) is identified as referential because although this is the first mention of the referent in the discourse, the addressee should be able to identify the intended referent on the page of the picture book in front of her.

Furthermore, our data shows that out of 43 instances of proximal demonstrative pronoun *zhe*, 27 occurrences (62.8%) are in focus and 16 occurrences (37.2%) are activated, as exemplified in (15) to (17). This is in contrast to Gundel et al.'s (1993) study, where only 2 instances (0.8%) of *zhe* were found and both were used for activated referents.

(15) **Zhe**   shi   shenme   zi   a?   [IN FOCUS]
     DEM   is   WH   character   PRT
     'What's this character?'

(16) **Zhe**   quandou   shi   ren.   [IN FOCUS]
     DEM   all   is   people
     'These are all people.'

(17) Hai   you   xiaogou   ye   zai   **zhe**.   [ACTIVATED]
     still   EXT   dog   also   at   DEM
     'There is also a dog here.'

In (15) and (16), the targeted referring expressions are coded as in focus because the speaker was pointing at the referents in the picture book when she spoke. In (17), the referring expression is only activated but not in focus because although the speaker has identified the location of the dog in the preceding utterance, the location is not the topic or focus of the preceding utterance, and the speaker was not making any gesture when she uttered the sentence.

Additionally, the frequencies of the proximal demonstrative determiner *zhe* in our data that are coded as in focus (12 occurrences) and activated (13 occurrences) are almost the same (48% vs. 52%), while the frequency of *zhe*-N used for activated referents (12 occurrences, 30.8%) in Gundel et al.'s (1993) study is more than twice as high as that of the proximal demonstrative

determiners used for in focus referents (26 occurrences, 66.7%). Examples of referring expressions with the proximal demonstrative determiner *zhe* found in our data are provided in (18) and (19).

(18) **Zhe-xie-xiaopengyou** dou bei [IN FOCUS]
DEM-PL-child *all* *BEI*
ta-men nong zou le.
3SG-PL make go PRT
'These children were all taken away by them.'

(19) **Zhe-ge-xiao-nühai** zai yi-ge chuan-shang. [ACTIVATED]
DEM-CL-small-girl PROG one-CL boat-above
'This little girl was on the boat.'

In (18), the referent of *zhe*-N is coded as in focus because not only has it been mentioned in the previous discourse, but the child was pointing at the story character in the picture book when she retold the story. In (19), the referent of *zhe*-N is considered as activated but not in focus because although the referent is discourse-new and the child did not make any gesture, the referent is still highly accessible to the addressee in the context.

Finally, according to Gundel et al. (1993), indefinite noun phrases *yi* N 'one N' in their study are most likely to be referential (17 occurrences, 89.5%), and bare nouns are most likely to be uniquely identifiable (49 occurrences, 47.1%). However, in our data, most of the indefinite noun phrases *yi*-N (19 occurrences, 86.4%) are merely type identifiable, as exemplified in (20), and most of the bare nouns are either activated (49 occurrences, 31.2%) or type identifiable (78 occurrences, 49.7%), as shown in (21) and (22).

(20) Ta shi **yi-ge-xiaotou** ma? [TYPE IDENTIFIABLE]
3SG is one-CL-thief Q
'Is he a thief?'

(21) Na-xie nan-ren kanjian [TYPE IDENTIFIABLE]
DEM-PL male-people see
nü-ren, ø dou yun-le.
female-people (RE) all faint-ASP
'Those men saw women, all fainted.'

(22) **Xiao-guaiwu-men** ye dao loushang [ACTIVATED]
small-monster-PL also towards upstairs
qu shai-le yifu
go hang-ASP cloth
'(The) little monsters also went upstairs and hanged (the) clothes.'

The targeted referring expressions in (20) and (21) are coded as type identifiable because they both have a generic interpretation and do not refer to any particular entities in the world. In (22), the referent of the targeted bare noun is considered to be activated because it has been mentioned by the speaker in the previous discourse.

Table 6. *Distribution of referring expressions in the recordings of 4yf01–05 from the Zhou Narrative corpus according to discourse functions.*

	Introduction	Re-introduction	Maintenance	Total
ø	1	5	24	30
PRO	3	29	39	71
zhe				
na				
zhe N	4	10	1	15
na N	3	4	4	11
yi N	12	3		15
N	38	56	11	105
QUANT-N	7	11	2	20
Total	68	118	81	**267**

In addition to the correlation between the referential forms and their cognitive statuses, this study also examines the distribution of the referring expressions found in our dataset according to their discourse functions. The results are presented in Table 6, which shows that out of a total of 424 tokens of referring expressions in our data, 267 occurrences are used to refer to the characters in the picture book. Within these 267 occurrences of referring expressions, 68 occurrences (25.5%) are used to introduce new entities to the discourse, while 199 occurrences (74.5%) are used to re-introduce or maintain reference to entities that have been introduced in previous discourse. Our results show that children are most likely to use bare nouns, noun phrases with the indefinite article *yi* 'one', and noun phrases modified by quantifiers or numerals for referent introduction and re-introduction, and to use overt pronouns and zero pronouns to maintain reference to story characters. These patterns generally follow the definiteness constraint on Chinese referring expressions predicted by the Givenness Hierarchy (Gundel et al., 1993) and the findings from Chen and Lei (2013). As mentioned, all referents in our dataset that are type identifiable and thus have an indefinite interpretation are indicated by either bare nouns, indefinite noun phrases or noun phrases modified by quantifiers/numerals, while all referents of overt or zero pronouns are either in focus or activated and thus have a definite interpretation. Table 7 presents the distribution of the discourse function of referring expressions found in our dataset according to their cognitive statuses.

Table 7. *Distribution of discourse functions of referring expressions in the recordings of 4yf01–05 from the Zhou Narrative corpus according to cognitive statuses*

	Focus	Activated	Familiar	Unique	Referential	Type	Total
Introduction	6	9		4		49	68
Re-introduction	15	89				14	118
Maintenance	78	2				1	81
Total	99	100		4		64	267

Table 7 presents the correlation between the function of referring expressions and the cognitive status of their referents. As expected, the distribution of discourse functions across cognitive statuses generally patterns with the distribution according to referential forms. Most of the referring expressions used for referent introduction indicate referents that are type identifiable (49 occurrences, 72.1%). In total, 104 occurrences of referring expressions used for referent re-introduction refer to referents that are either in focus or activated, accounting for 88.1% of all the referring expressions found in our dataset that are used for referent re-introduction. Furthermore, almost all the referring expressions used to maintain reference to discourse-old entities found in our data also refer to referents that are at least activated. Only one occurrence of referring expression in our data used for referent maintenance refers to type identifiable entities, as shown in (23). The targeted referring expression in (23) is coded as bare noun used to maintain reference to type identifiable referent because the speaker simply intended to assert that there are such three people in the world of the book, and the utterance in (23) is a repetition of the immediately preceding utterance.

(23) Hai shi you **san-ge-ren**. [MAINTENANCE; TYPE IDENTIFIABLE]
     still is EXT three-CL-people
     'Still there are three people.'

In addition to the cognitive status and discourse function, this study also investigates the distribution of referring expressions according to their grammatical roles. Based on the findings from the previous studies on Chinese referring expressions (e.g., Kuo, 2008; Nie, 2020; Shi, 1998; van Deemter et al., 2017), this study examines the referring expressions in subject and object positions not only in matrix clauses, but also in existential *you*-sentences, *ba*-sentences, and *bei*-sentences.[i] The results are presented in Table 8, which shows the distribution of referring expressions occurring in our data according to their grammatical roles. The raw frequencies are indicated in the Freq. column and the percentages are indicated in parentheses.

Table 8. *Distribution of the referring expressions in the recordings of 4yf01–05 from the Zhou Narrative corpus according to grammatical roles*

Rank	Position	Freq. (%)
1	SUB	208 (52.5%)
2	OBJ	119 (30.1%)
3	OBJ in EXT	42 (10.6%)
4	SUB in EXT	12 (3.0%)
5	OBJ in BA	11 (2.8%)
6	OBJ in BEI	2 (0.5%)
7	SUB in BA	1 (0.3%)
8	SUB in BEI	1 (0.3%)
Total		396 (100%)

Table 8 shows that more than 80% of the referring expressions in our data occur in matrix clauses, and the frequency of referring expressions in the subject (SUB) position (208 times, 52.5%) is much higher than that in the object (OBJ) position (119 times, 30.1%). Furthermore, 42 tokens (10.6%) of referring expressions are found in the subject position of existential *you*-sentences (EXT), and 12 (3.0%) tokens are found in the object position, representing 13.6% of all occurrences of referring expressions in our data. Eleven occurrences of referring expressions are found in the object position of *ba*-sentences (BA), while only one referring expression is found in the subject position of *ba*-sentences. Only three instances of referring expressions are found in *bei*-sentences (BEI): two in the object position and one in the subject position. To better understand the role of sentence structure in the choice of the various forms of referring expressions for different discourse functions, we further investigate the correlation between the form, grammatical role, and discourse function of the referring expressions found in our data. The results are presented in the following tables.

Table 9 presents the distribution of different forms of referring expressions organized by their grammatical roles. The distributional pattern of referring expressions in our data generally follows the universal given-before-new order of information structure. Zero pronouns, overt pronouns, and demonstratives are more likely to occur in the subject position of matrix clauses than in the object position, and these are the forms which only refer to referents that are at least activated in our data, while no such preference exists for noun phrases with the indefinite article *yi* 'one', bare nouns, or noun phrases modified by quantifiers/numerals. In addition to matrix clauses, the sentence structure of existential *you*-sentences also seems to play a role in the choice of forms of referring expressions. In our dataset, only zero pronouns are found to occur in the subject position of existential *you*-sentences, while almost all objects of *you*-sentences are either indefinite noun phrases with the determiner *yi* 'one', bare nouns, or noun phrases modified by quantifiers/numerals.

Table 9. *Distribution of referring expressions in the recordings of 4yf01–05 from the Zhou Narrative corpus according to grammatical roles.*

	SUB	OBJ	SUB (EXT)	OBJ (EXT)	SUB (BA)	OBJ (BA)	SUB (BEI)	OBJ (BEI)	Total
ø	35	4	12		1	1			53
PRO	59	15				2	1	2	79
zhe	34	6				1			41
na		1							1
zhe N	14	9		1					24
na N	6	5				1			12
yi N	4	12		4					20
N	48	58		26		6			138
QUANT-N	8	9		11					28
Total	208	119	12	42	1	11	1	2	396

Table 10. *Distribution of discourse functions of referring expressions in the recordings of 4yf01–05 from the Zhou Narrative corpus according to grammatical roles.*

	SUB	OBJ	SUB (EXT)	OBJ (EXT)	SUB (BA)	OBJ (BA)	SUB (BEI)	OBJ (BEI)	Total
Introduction	25	26		13					64
Re-introduction	60	26		18		5		1	110
Maintenance	62	11		1	1	3	1	1	80
Total	147	63		32	1	8	1	2	254

Table 10 shows the distribution of the discourse function of referring expressions found in our dataset organized by their grammatical roles. As expected by the universal given-before-new principle, referring expressions used to re-introduce or maintain reference to discourse-old entities are much more likely to occur in the subject position of matrix clauses than in the object position. In addition to matrix clauses, referring expressions occurring in the object position of existential *you*-sentences are rarely used for referent maintenance function. However, Table 10 also shows some patterns that do not seem to follow the general definiteness constraints on referring expressions. For example, referring expressions used for referent introduction show no preference in terms of sentence structure in matrix clauses. Expressions occurring in the object position of existential *you*-sentences are often used to re-introduce entities that have been mentioned in previous discourse. To account for these patterns, further qualitative analyses of the data are required.

5   Conclusion

Referring expressions are pervasive and play a significant role in daily communication across languages. Based on the data from the Zhou Narratives

corpus (Li & Zhou, 2011), this paper investigates the use of referring expressions in the spoken narratives of five monolingual Chinese-speaking children at age four. By adopting Gundel et al.'s (1993) cognitive approach and the Givenness Hierarchy, this research provides a preliminary analysis of the quantitative results obtained from the corpus study. Our findings indicate that the correlations between the forms of Chinese referring expressions produced by 4-year-old children and their cognitive statuses and discourse functions align with the predictions of the Givenness Hierarchy. These results are consistent with Gundel et al.'s (1993) study, which was based on data from adult Chinese speakers. Overall, for both children and adult Chinese speakers, bare nouns and overt pronouns are used most frequently, whereas *yi*-N ('one-N'), *na*-N ('that-N'), and the distal demonstrative pronoun *na* are all relatively rare in the data.

The study also reveals some noteworthy differences between Chinese monolingual adults and children in their preferred referential forms for certain cognitive statuses and discourse functions. Specifically, children are more likely to utilize the proximal demonstrative pronoun *zhe* compared to adults, while adults use *zhe*-N 'this-N' more frequently than children. Regarding discourse functions: (1) adults predominantly use zero pronouns for in focus referents, while children use them for both in focus and activated referents; (2) overt pronouns are only used for in focus referents for adults; however, for children, they are used for both in focus and activated referents; (3) adults prefer to use the proximal demonstrative determiner *zhe* for in focus referents over activated ones; in contrast, children employ *zhe*-N 'this-N' for both in focus and activated referents with equal frequency; (4) for adults, indefinite noun phrases *yi*-N 'one-N' and bare nouns are most likely to be referential and uniquely identifiable, respectively; however, for children, *yi*-N 'one-N' is at most type identifiable, and bare nouns are either type identifiable or activated.

In conclusion, our findings suggest that the development of referential appropriateness in discourse is a gradual process for children. This study provides the first systematic investigation of the referential forms and their cognitive statuses and discourse functions in the discourse of Chinese monolingual preschool children. It contributes to a deeper understanding of the developmental trajectory of referential appropriateness in children's discourse and has implications for educational and developmental language interventions for preschool children's discourse development.

References

Ariel, M. (1991). The function of accessibility in a theory of grammar. *Journal of Pragmatics*, *16*, 443–463. https://doi.org/10.1016/0378-2166(91)90136-L

Carle, E. (1969). *The very hungry caterpillar*. Philomel Books.

Chao, Y. R. (2011). *A grammar of spoken language*. The Commercial Press. (Original work published 1968)

Chen, L., & Lei, J. (2013). The production of referring expressions in oral narratives of Chinese-English bilingual speakers and monolingual peers. *Child Language Teaching and Therapy*, *29*(1), 41–55. https://doi.org/10.1177/0265659012459527

Chen, L., & Pan, N. (2009). Development of English referring expressions in the narratives of Chinese English bilinguals. *Bilingualism*, *12*(4), 429–445. https://doi.org/10.1017/S1366728909990216

Fuller, J. W., & Gundel, J. K. (1987). Topic-prominence in interlanguage. *Language Learning*, *37*(1), 1–18. https://doi.org/10.1111/j.1467-1770.1968.tb01310.x

Givon, T. (1983). Topic continuity in discourse: An introduction. In: Givon, T. (Ed.), *Topic continuity in discourse: A quantitative cross-language study* (pp. 1–43). John Benjamins Publishing Company.

Gundel, J. K., Hedberg, N., & Zacharsk, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*(2), 274–307. https://doi.org/10.2307/416535

Kuo, J. Y. (2008). A pragmatic approach to the interpretations of Mandarin bare nouns. *Journal of Pragmatics*, *40*(6), 1082–1102. https://doi.org/10.1016/j.pragma.2008.03.004

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. University of California Press.

Li, L. H., & Zhou, J. (2011). *Preschool children's development of reading comprehension of picture storybook: From a perspective of multimodel meaning making*. East China Normal University.

Mayer, M. (1969). *Frog, Where Are You?* Dial Press.

Nie, X. (2020). *Information structure, discourse and distribution of referring expressions in Chinese*. The University of Arizona.

Serratrice, L. (2007). Referential cohesion in the narratives of bilingual English–Italian children and monolingual peers. *Journal of Pragmatics 39*, 1058–1087. https://doi.org/10.1016/j.pragma.2006.10.001

Séchan, E. (Director). (1959). *The golden fish* [Film]. Columbia Pictures.

Shi, Y. (1998). *Referring expressions in Chinese and English discourse*. Ball State University.

Ungerer, T. (1962). *The three robbers*. Atheneum.

van Deemter, K., Sun, L., Sybesma, R., Li, X., Bo, C., & Yang, M. (2017). Investigating the content and form of referring expressions in Mandarin: Introducing the Mtuna corpus, 213–217. https://doi.org/10.18653/v1/W17-3532

Yang, C. L., Gordon, P. C., Hendrick, R., & Wu, J. T. (1999). Comprehension of referring expressions in Chinese. *Language and Cognitive Processes*, *14*(5–6), 715–743. https://doi.org/10.1080/016909699386248

ⁱ The *ba*-sentence, *bei*-sentence, and *you*-sentence are three frequently discussed constructions in Mandarin grammar. According to (Li & Thompson, 1981), a *ba*-sentence typically has a structure of subject + *ba* + direct object + verb, as exemplified in (24). The NP that follows *ba* is generally definite or generic. The example in (24) is appropriate when the speaker assumes that the addressee knows what chair is being referred to.

(24) Wo   jintian   ba   san   ben   shu   dou   mai   le.
      1SG   today   BA   three   CL   book   all   sell   ASP
      'I sold all three books today.'   (Li & Thompson, 1981, p. 465)

The *bei*-sentence is the default construction for passive sentences in Mandarin and generally follows a linear structure of NP1 *bei* NP2 verb, as illustrated in (25). The first NP in example (25) *ta* '3SG' functions as the direct object affected by the action of the verb *ma* 'scold'. *Bei* can be considered a passive coverb, while NP2 *jiejie* 'older sister' is the agent of the action – the one who did the scolding.

(25) Ta   bei   jiejie   ma   le.
      3SG   BEI   older sister   scold   ASP
      'S/he was scolded by (his/her) older sister.'   (Li & Thompson, 1981, p. 492)

Lastly, the existential *you*-sentence in Mandarin is used to signal the existence of the referent of a NP at some place – locus. It can manifest in either of two structures: *you* + NP + *zai* 'at' + locus + (VP) or (*zai*) + locus + *you* + NP + (VP), as shown in (26) and (27).

(26) You   yi   zhi   gou   zai   yuanzi   li.
      EXIST   one   CL   dog   at   yard   In
      'There's a dog in the yard.'   (Li & Thompson, 1981, p. 511)

(27) (Zai)   yuanzi   li   you   yi   zhi   gou.
      at   yard   in   EXIST   one   CL   dog
      'There's a dog in the yard.'   (Li & Thompson, 1981, p. 510)