

1 Introduction

1.1 Functions of *de*

Both the semantic and syntactic characteristics of *de* have been a source of contention in the literature for well over forty years now. One reason for this disgruntlement may be that *de* wears many hats in Modern Mandarin and can combine with a large number of syntactic objects. When the target of modification is a noun phrase (NP), for instance, *de* may appear adjacent to an adjectival phrase (1), a possessive phrase (2), a prepositional phrase (PP) (3), or an inflectional phrase (i.e. as a complementizer) (4).

- (1) 胖 的 孩子
pang de haizi
fat DE child
'fat child'
- (2) 张三 的 卡车
zhangsan de kache
Zhangsan DE car
'Zhangsan's car'
- (3) 电影院 的 旁边
dianyingyuan de pangbian
cinema DE next-to
'next to the cinema'
- (4) 李斯 昨天 买 的 书
lisi zuotian mai de shu
Lisi yesterday buy DE book
'The book that Lisi bought yesterday'
- (5) 我 是 从来 不 抽 烟 的
wo shi conglai bu chou yan de
1sg be ever NEG inhale smoke DE
'(It is the case that) I never smoked' (example from Paul (2010))

Furthermore, there is a phonetically identical element that allows modification of verb phrases, but is distinguished orthographically. Examples (6) and (7) illustrate this use.

- (6) 跑 得 快
pao de kuai
run DE fast
'to run fast'

- (7) 紧张 地 问
 jinzhang de wen
 nervous DE ask
 'to ask nervously'

1.2 Theoretical Accounts

Given its vast range of applications within a sentence, a number of questions remain as to the exact characterization of *de* in Mandarin. Could it be, as proposed by Li and Thompson (1981), that there are several (functionally) different *des* in Mandarin, each one roughly corresponding to one of the above examples? Or are there only two types of *de*, a head-initial and a head-final complementizer, as put forth by Cheng (1986)? More recently, den Dikken and Singheprecha (2004) proposed a unified 'linker' analysis, whereby *de* is used to generate complex NPs by embedding a predicate inside the NP, and can therefore be combined with any predicative object in order to modify a nominal phrase. In more concrete terms, they assert that the modifying XP originates in the predicative position of an utterance with the NP as its subject, and is subsequently raised to SpecNP and linked to the head noun via insertion of *de*. Paul (2010) refutes this, however, on the basis that this interpretation incorrectly predicts that certain XPs can precede *de* to modify a NP. Furthermore, he argues, this analysis does not accurately reflect the canonical SVO word order of Mandarin. He proposes instead that as a non-root C(omplementizer) in constructions such as (4)

The absence or presence of *de* in a given utterance is subject to certain semantic considerations of both the modifier and the modified NP. For instance, *de* can be omitted before kinship terms when the modifying DP is a pronoun, as in (8). It can also be omitted in structures such as (9), albeit with an interpretative shift.

- (8) 她 (的) 哥哥
 3sgFEM (de) gege
 she (DE) older-brother
 'her older brother'
- (9) 木头 (的) 桌子
 mutou (de) zhuozi
 wood (DE) table
 'a wooden table' (example from Paul (2010))

2 Acquisition of *de*

2.1 Past findings

Despite its widespread use in the adult grammar and ongoing debate as to its typological characteristics, nobody to date has examined the development

of *de* in the grammar of Mandarin-speaking children. Like many linguistic phenomena, acquisition data may provide us with deep insight into the role that this particle plays in the grammar.

The acquisition path for *de* is also interesting for a number of theoretical reasons. Other morphemes with similar functions cross-linguistically, such as the Saxon genitive (-s) in English, are acquired only after three years of age (citation Radford and Galasso, 1998), similar to embedded relative clauses (cf. Berman 1997; Crain, McKee and Emiliani, 1990; others). However, both of these structures in Mandarin are marked using the *de* morpheme, along with a number of other modifications which are represented using distinct morphosyntax in many Indoeuropean languages; perhaps the wider applicability of *de* relative to structures with similar semantic purposes would shorten the acquisition timeline. If this is the case, it would raise questions regarding the role of both function and frequency on the acquisition of specific morphemes.

2.2 This study

- (10) 红色 的 给 我 吧
hongse de gei wo ba
red DE give 1sg IMP
'Give me the red one'

The present study endeavors to answer several questions regarding the normal acquisition of the *de* morpheme in Mandarin Chinese. Most importantly, we hope to determine at what point productive use of *de* emerges, and in what contexts it appears in early child speech. With regard to the syntactic and semantic context, a few qualitative characteristics present themselves for further inspection, namely:

Preceding item What types of items immediately precede *de* in child speech? As demonstrated in the previous sections, *de* can occur adjacent to a number of syntactic objects, including both nominals and predicates. If these different combinatorial possibilities represent distinct syntactic structures in the adult grammar, it is possible that the child learner would acquire each one at a different rate. On the other hand, if all of the possible uses of *de* can be subsumed under a single notion of 'linking element,' e.g. as proposed by den Dikken and Singheprecha, then we would expect to see all of the possible combinations emerge at the same time.

Head type Canonically, *de* is used for complex nominals such as (1), (2), and (4). However, it can also appear in a sentence-final position in contexts like (5), where it combines with copular *shi* to emphasize the preceding tense phrase. Furthermore, due to Mandarin's status as a radical pro-drop or 'topic-drop' language, both subject and object nominal phrases modified by *de* can be omitted if the referent is previously established in the discourse, e.g. (10). In these cases, it would appear that the 'head' modified

by *de* is a predicate such as a VP or an adjective, when in fact it is an unpronounced NP.

To that end, an analysis of several corpora was conducted in order to determine roughly when and how *de* develops in the grammar of Mandarin-speaking children.

3 Method

A corpus search was conducted using corpora available via the CHILDES database (MacWhinney, 2000). A complete list of the corpora used, along with their basic attributes, can be seen in Table (1). Corpora were included if the youngest age in the dataset was no greater than three years (36 months) old, providing a total of six different datasets to be considered in this analysis. Although a number of the corpora included older children, the maximum age included in this search was 48 months.

| Corpus name | Author | Age range (mo) | Type | Collection frequency |
|-------------|--------------------------------|----------------|-----------------|----------------------|
| Erbaugh | Mary Erbaugh and Linhui Li | 24-45 | Longitudinal | Monthly |
| Tong | Xiangjun Deng and Virginia Yip | 19-40 | Longitudinal | Monthly |
| Zhou 3 | Jing Zhou | 8-65 | Longitudinal | Monthly (20-30m) |
| Zhou 1 | Jing Zhou | 14-32 | Cross-sectional | Single session |
| Zhou 2 | Jing Zhou | 36-72 | Cross-sectional | Single session |
| LiZhou | Linhui Li and Jing Zhou | 36-72 | Cross-sectional | Single session |

Table 1: Basic information about the corpora used in the search.

The Erbaugh, Tong, and Zhou 3 corpora contain data from longitudinal studies of one or more Mandarin-acquiring children. The Erbaugh and Tong studies collected data approximately once a month for the duration of the study; the Zhou 3 authors collected data each month when the target child was between 20 and 30 months old, and then decreased the frequency to about every other month when the child was between 30 and 50 months old. After that, the frequency decreased further; however, the later data is not included in the present analysis.

The Zhou 1, Zhou 2, and Li Zhou corpora are cross-sectional studies conducted with a number of Mandarin-speaking children at kindergartens across Mainland China. The data in each of these corpora was collected during a single recording session. Thus, unlike the first three corpora, these data do not reflect the acquisition path for a single child, but are intended to reveal more general trends across age groups. The Zhou 1 corpus contains data from groups of 14, 20, 26, and 32 month old children. Likewise, the Zhou 2 corpus divides its data into seven age groups: 36, 42, 48, 54, 60, 66, and 72 months old. The Li Zhou corpus contains data for children between ages three and six. In all

cases, however, only data for children less than or equal to 48 months old were included.

The analysis was run in Python (version 3.8.7), using Natural Language Toolkit (NLTK, version 3.5)’s built-in corpus reader functionality for CHILDES. XML files for each database were downloaded manually from the TalkBank website and searched using a program written in Python (see 6 for more detailed information). Further analysis was carried out using the NumPy, Pandas (1.2.1), and Matplotlib packages.

Two common expressions containing *de* were excluded from the search criteria. The first, *de shi hou* 的时候, translates roughly to English ‘while’ and occurs in IP-final position. The other excluded expression, *de hua* 的话, also occurs IP-finally and can be translated as ‘if’ or ‘in the case that...’. Because the primary purpose of this analysis is to determine when and how Mandarin acquiring children begin to use *de* productively, their ability to use it in a set phrase such as these was not considered relevant to the present research question.

More than 3,800 child utterances resulted from the initial search, excluding the two aforementioned expressions. From this general dataset, I created several subsets of data to be analyzed individually and compared with one another. Each of these subsets will be briefly characterized below.

3.1 Data subsets

The Erbaugh (citation), Tong (citation), Zhou 1 (citation), and Zhou 3 (citation) corpora were each searched and analyzed individually. The reason for this is that each corpus collected its samples beginning at slightly different ages, and with differing regularity, thus a dataset which combines all results for these corpora might falsely inflate the number of *de* uses for overlapping ages, while underrepresenting the number of uses for age ranges that were only collected by one or two of the studies. The Zhou 3 corpus, for instance, began collecting data when the child was eight months old, whereas the Erbaugh corpus began collecting data at 24 months. Thus, there is less data overall for some ages compared with others, which would misconstrue the raw numbers. There were 572 total *de* utterances contained in the Tong corpus; 211 in the Zhou 1 corpus; 372 in the Zhou 3 corpus; and 1,969 in the Erbaugh corpus. The fact that each study had its own sampling frequency and length should be kept in mind when comparing results to other corpora in this analysis.

There is an overlap between 24 and 30 months old during which period all the Tong, Erbaugh, and Zhou 3 studies collected data on a monthly basis, albeit for different lengths of time. Since the representativeness of the data would be more comparable for this period, and the ages targeted tend to constitute a period of considerable progress in the acquisition of syntax, an additional subset of data consisting of utterances from all three corpora was created. The total number of utterances in this dataset was 582.

Finally, an additional dataset was created containing utterances from children between 36 and 48 months old. These included data from the aforementioned corpora, as well as two additional corpora that can be viewed in Table

(1). This was a particularly robust data set, containing 2,054 total utterances.

3.2 Data omissions

Several types of MOR (part of speech) tags appeared in the initial search for items immediately following *de* that were either incorrectly coded, overlapped with existing tags, or otherwise warranted removal from the dataset. Some results were also collapsed into single categories for ease of analysis, for instance, all verbs with the exception of *shi* were tagged as VP (*shi* was tagged as a copular verb). When *de* was followed by an item tagged as a 'sentence-final particle (SFP),' the utterance was assigned to the 'sentence final' category. Likewise, if *de* was followed by the aspectual marker *le* or a conjunction such as *yinwei* 'because', the utterance was tagged as sentence-final. An 'other' tag was created for infrequent values that occurred only a handful of times in the data. This included items tagged as 'possessive', 'chi', and 'classifier,' among a few others.

4 Results

4.1 Earliest uses

(11) Zhou1/cs14b.xml (1;02)

要 的
yao de
need DE
'need/should'

The youngest age at which *de* was used by a child was fourteen months old (see (11)); however, this was limited to a single utterance in the Zhou 1 corpus (citation). *De* did not appear again in the same corpus (or any other corpora, for that matter) until twenty months old, at which point it is used twelve times by the same child, and once each by two children from the Tong and Zhou 3 corpora. In eight of the fourteen total utterances containing *de* at 20 months old, *de* occurs in a sentence-final position. In the remaining six non-SF utterances, the head items (i.e. the word which is modified by *de*) consist of four nouns, one adjective, and one verb.

From the onset of the two-word utterance phase,

5 Discussion

6 Appendix