

1 Introduction

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'

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 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, Paul (2010) has suggested that

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 as *de* 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, whereas they are wholly dissimilar in English and many other Indo-European languages; perhaps the greater flexibility with which it can be used in the target language would shorten the acquisition timeline. If this is the case, it would raise a number of questions regarding the role of both function and frequency on the acquisition of specific morphemes.

2.2 This study

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

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. 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

| 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 (20-30 months) |
| 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.

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, this 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 14, 20, 26, and 32 month old children. The Zhou 2 corpus divides its data into seven age groups: 36, 42, 48, 54, 60, 66, and 72 months old; however, only the first three age groups were included in this analysis.

The analysis was run in Python, using Natural Language Toolkit (NLTK, version 3.5)’s built-in corpus reader functionality for CHILDES. XML files for each database were downloaded from the TalkBank website and searched using a program written in Python (see 6 for the complete code). 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, 的时候, translates roughly to English ‘while’ and occurs in IP-final position. The other excluded expression, 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 6,000 child utterances resulted from the initial search, excluding the two aforementioned expressions. From this general dataset, I create several subsets of data in order to give a more accurate reflection of the distributions. 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 for the only data for e.g. 8 to 15 months comes only from Zhou 3, which could misconstrue the raw numbers. There were 587 total *de* utterances contained in the Tong corpus; 566 in the Zhou 3 corpus; and 1,987 in the Erbaugh corpus. The comparative extensiveness of the Erbaugh corpus 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 three 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, an additional subset of data consisting of utterances from all three corpora was created. The total number of utterances in this dataset was 592. Finally, an additional dataset was created containing utterances from children between 36 and 48 months old. These included data from the Erbaugh, Tong, and Zhou 3 corpora, as well as additional corpora that can be viewed in Table (1). This was a particularly robust data set, as many of the Mandarin corpora in CHILDES began their study around the child(ren)’s third birthday. The total number of utterances contained in this subset was 2,361 utterances.

4 Results

4.1 Earliest uses

(8) 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 (8)); 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