The Error Analysis of De Structures: A Corpus-based Research

Shuang Ding[1,a,*]

[1]College of Humanities and Law, South China Agricultural University, WuShan Road, TianHe District, Guangzhou, China

a. 201911310502@stu.scau.edu.cn

*corresponding author

Abstract: This article analyzes the errors in de structures produced by learners studying Chinese as a second language based on the Global Chinese Interlanguage Corpus. The research is hoped to open up a new vision for error analysis of de structures with methods of corpus linguistics and provide some empirical evidence that can be used to adjust the teaching of Chinese grammar to make it more problem-oriented. The article begins with a brief review of relevant studies and an explicit definition of the de structure. Data used in the article all come from the Global Chinese Interlanguage Corpus, which has been sifted carefully to make them better adapted to the research objective. Errors in de structures are categorized into five different types based on the ways surface structures are altered. The characteristics of each error type are given by the description and correction of example sentences. Two mechanisms are given as the possible explanation for the roots of errors in de structures: L1 transfer and overgeneralization. The contrastive analysis between Korean, English and Chinese explains the occurrence of serval types of errors, and some erroneous sentences created by CSL learners also confirm the existence of overgeneralization.

Keywords: error analysis, corpus linguistics, interlanguage

1. Introduction

The de structure is one of the fundamental linguistic structures in the contemporary Chinese language. This widely used structure is constructed by the combination of content constituents and the particle de (de 的). As a nominal phrase, de structure mainly functions as a subject or object in sentences. Due to its complication in both syntactic and semantic levels, learners studying Chinese as a second language (CSL) usually encounter obstacles when applying de structures, and errors yielded in this process provide a viable path to analyze how de structures are acquired and, more broadly, the mechanism of second language acquisition.

So far, the error analysis of de structures has attracted great attention, and researchers have conducted many studies. Most of these studies used questionnaires to examine CSL learners' linguistic performance on de structures or focused on a small number of learners to conduct a diachronic case study. Although both methods have proved effective and produced abundant results, more perspectives are still needed, given the complexity of error analysis in Second Language

Acquisition (L2 acquisition). Against such background, applying methods of corpus linguistics is essential as it opens up a novel vision for error analysis of de structures.

The article includes three parts. Chapter 1 begins with the explicit definition of the de structure on which the following collection of data is based. A brief introduction to the Global Chinese Interlanguage Corpus, the process of data collection and sifting and the rudimentary classification of data is also involved in this section. Chapter 2 moves on to the categorization of errors. Based on the taxonomy developed by Dulay, Burt and Krashen, and James, all these errors are divided into five types: omission, overinclusion, misselection, misordering and blend [1,2]. The characteristics of each error type are given by the description and correction of example sentences. Chapter 3 gives two possible explanations for the roots of errors in de structures: first language transfer (L1 transfer) and overgeneralization. The contrastive analysis between Korean, English and Chinese explains the occurrence of serval types of errors, and some erroneous sentences created by CSL learners also confirm the existence of overgeneralization.

2. The Uses of De Structures

2.1. The Delimitation of De Structures

The specific study on de structures started with the publication of On De, where Zhu analyzed the functions of de with methods of structural linguistics [3]. Since then, the de structure has been one of the important yet contentious topics in the studies of Chinese grammar, with a large number of debates revolving around it. Some scholars define the de structure as an elliptical form of the subordinative phrase, from which the head (i.e. center of an endocentric construction) is omitted [4]. In the contemporary Chinese language, the modifier and the head are usually connected by the word de to form a subordinative phrase as in 红的花 (hongdehua, 'red flower'), and de works as a marker of modification in such context. By omitting the head of the subordinative phrase, the structure remains its original meaning, and one could restore the omitted part if enough information is given. For instance, in the sentence 比起白的花，她更喜欢红的。(biqi baidehua ta gengxihuan hongde, 'Compared to white flowers, she prefers red ones.'), head in the second subordinative is omitted and thus constitutes a de structure. De actually retains its status as the marker of modification in this sense. Another popular viewpoint is to decouple the de structure from the subordinative phrase, defining it as a self-standing construction formed by the combination of content constituents and the nominalizer de [3]. For instance, the adjective 热 (re, 'hot') can convert to a nominal construction by attaching the nominalizer de to it, resulting in 热的 (rede, 'hot things').

To facilitate the research, de is marked as a nominalizer in the interlinear glosses, but theories from both viewpoints are included. The article invokes the argument proposed by Huang & Liao as an explicit definition of the de structure: de structures are constructed by the particle de attaching after content words or phrases, referring to a person or thing [5]. The de structure is a nominal phrase [5]. It can work as a subject or object in a sentence [5].

2.2. The Error Rate of De Structures

Data used in the article come from the Global Chinese Interlanguage Corpus (GCIC), the largest Chinese Interlanguage Corpus for now, with 85,278,546 words in total which includes written, spoken, and video corpus of CSL learners from 111 countries and regions [6]. Considering its large scale and abundant types, the situation of how de structures were used by CSL learners can be fully demonstrated. Enter the search system of GCIC; Select "retrieve by normal strings"; Type the keyword "de"; Search without any other restriction, and 241,517 results are found. After random sampling, 2,000 samples are selected out of 241,517 results. Since the object of the research is the de

structure, the following samples are not taken into account: de used as a marker of modification, de used as an aspectual particle, de used as a modal particle, de used as a morpheme in words, as in the adverb 真的(zhende, 'really'). After excluding all irrelevant samples, there remain 155 valid samples, among which 87 samples are incorrect in the use of de structures, with an error rate of 56%.

As noted before, de structures are constructed by the combination of content constituents and the particle de, which can be formulated as X+de, with X representing content words or phrases [7]. Therefore, 155 samples could be categorized depending on the feature of X.

Table 1: Categories of all samples based on the feature of X.

Grammatical categories of X		Frequency	Relative Frequency
Word	Noun	12	8%
	Pronoun	11	7%
	Verb	20	13%
	Adjective	24	16%
	Adverbial	14	9%
	Total	81	52%
Phrase	Coordinate Phrase	3	2%
	Subordinative Phrase	33	21%
	Verb-Object Phrase	15	10%
	Subject-Predicate Phrase	21	14%
	Pivotal Phrase	2	1%
	Total	74	48%

Table 2: Categories of erroneous samples based on the feature of X.

Grammatical categories of X		Frequency	Relative Frequency (of erroneous samples)
Word	Noun	6	7%
	Pronoun	1	1%
	Verb	16	18%
	Adjective	13	15%
	Adverbial	14	16%
	Total	50	57%
Phrase	Coordinate Phrase	2	2%
	Subordinative Phrase	16	18%
	Verb-Object Phrase	11	13%
	Subject-Predicate Phrase	8	9%
	Pivotal Phrase	0	0%
	Total	37	43%

Table 1 categorized de structures based on the feature of X. The results indicate that the relative frequency of subordinative phrases, adjectives and subject-predicate phrases is 21%, 16% and 14%, respectively, which are the three most frequent forms. Table 2 demonstrates the breakdown of errors in de structures, and the results show that the three most problematic forms are subordinative phrase

(18%), verb (18%) and adverbial (16%), not completely proportional to the relative frequency in table 1. It is noteworthy that the error rate of adverbial de structures is 100%, and this closely relates to the feature of de, which will be discussed in more detail later.

3. The Error Types of De Structures

Dulay, Burt and Krashen developed a descriptive error taxonomy dividing errors into four types based on a comparison of the erroneous forms the learner used with the intended correct forms [1]. James modified Dulay et al.'s taxonomy by coalescing and relabeling the categories and proposed a new taxonomy with five error types instead of four: omission, overinclusion, misselection, misordering and blend [2]. The taxonomy used in the article is the version of James'. For better interpretation, interlinear glosses are added under original sentences, which follow the Leipzig Glossing Rules, incorporating some notions proposed by Chen et al. to make these rules better adapted to Chinese [8].

3.1. Omission

Omission errors are characterized by the absence of an item that must appear in a well-formed structure, either a content word or a grammatical morpheme [1]. Omission errors in de structures comprise three subtypes: the omission of de, the omission of the head, and the omission of other constituents.

3.1.1. Omission of the De

(1)幸运	是,	中国	朋友	来	接	我们。	
xingyun	shi,	zhongguo	pengyou	lai	jie	women	
lucky	is	China	friend	come	pick	us	

'The lucky thing is that Chinese friends come and pick us up.'

(2)因为	她的行李	跟	那个	人	一样。	
yinwei	ta-de-xingli	gen	na-ge	ren	yiyang	
because	she-NM.POSS-baggage	with	that-one	person	same	(NM-Noun Marker)

'Because her baggage is the same as that person's.'

According to Dulay, Burt and Krashen, grammatical morphemes without lexical meanings are more prone to be omitted by L2 learners compared to content morphemes [1]. The sentences listed above are both erroneous due to the omission of the grammatical morpheme de. Inserting de into the right positions will convert them into the correct forms. Sentence (1) is ungrammatical as a nominal constituent is needed here to act as the subject in the main clause. Attaching the nominalizer de after the adjective 幸运 (xingyun, 'lucky') will convert it into the de structure 幸运的 (xingyunde, 'lucky thing'). The situation of the sentence (2) is slightly different: it is grammatically acceptable. However, when it comes to the semantic level, sentence (2) demonstrates some incongruity. If accepting it as a valid sentence, the literal meaning of the sentence (2) will be because her baggage is the same as that person, contrasting her baggage and that person, which is obviously not the original intention of the writer. The de structure 那个人的 (nagerende 'that person's') should be applied here to refer to that person's (baggage).

3.1.2. Omission of the Head

(3)我	发现	巴西木	总是	那副	绿油油的,	一直	那个	样子。
wo	faxian	baximu	zong-shi	na-fu	Lüyouyou-de,	yizhi	na-ge	yangzi
I	find	brazilwood	always-is	that-pair	brightly.green- NMLZ	always	that-one	appearance

'I found that the brazilwood was always brightly green and always remained the same appearance.'

Generally, subordinative phrases can convert to de structures through the ellipsis of the head. However, such transformations also have many restrictions, which sentence (3) ignored. Lü pointed out that when the head of a subordinative phrase is an abstract noun or an appellation, the ellipsis is forbidden [9]. In sentence (3), 样子 (yangzi, 'appearance') is a member of abstract nouns and cannot be omitted to form a de structure.

3.1.3. Omission of Other Constituents

(4)生活里　　有　　许许多多的故事，　　有的　　　开心，　　有的　　　生气。
shenghuo-li　you　xuxuduoduo-de-gushi,　you-de　kaixin,　you-de　shengqi
life-in　　　have　many-NM-story　　　have-NMLZ　happy　have-NMLZ　angry
'There are many stories in my life; some are happy, some are angry.'

(5)尊敬　　　老人，　　这个　　　是　　那时候　　　体会的。
zunjing　lao-ren,　zhe-ge　shi　na-shihou　tihui-de
respect　old-man　this-one　is　that-time　experience-NMLZ
'Respecting the elder, this is what I experienced at that time.'

In sentence (4), the adjectives 开心 (kaixing, 'happy'), 生气 (shengqi, 'angry') and 委屈 (weiqu, 'feel wronged') require the NPs modified by them to have the semantic feature of [+human], but de structures 有的 (youde, 'some') here refer to 有的故事 (youde gushi, 'some story'), which violates the S-selection of aforementioned adjectives. In sentence (5), 体会 (tihui, 'experience') is an intransitive verb and has to be followed by a complement 到 (dao, 'to') when used as a de structure. The errors in these two sentences are actually matters of C-selection and S-selection, but this error type is relatively small in number and has little to do with the topic of the article. Therefore, they are classified in this general category and will not be discussed separately.

3.2. Overinclusion

Overinclusion errors, opposite to omissions, are characterized by the presence of an item which must not appear in a well-formed utterance [1]. Based on the samples from CGIC, overinclusion errors in de structures are relatively simple, mainly related to the incorrect addition of de itself.

(6)我　希望　　做的　　　　老师。
Wo　xiwang　zuo-de　laoshi
I　hope　be- NMLZ　teacher
'I hope to be a teacher.'

(7)我　这次　　也　　表示　坏的样子，　　因为　　拉不下　　脸的。
Wo　zhe-ci　ye　biaoshi　huai-de-yangzi,　zhangfu　la-bu-xia　lian-de
I　this-time　also　express　bad-NM-appearance　because　pull-not-down　face- NMLZ
'I also acted unfriendly this time because I was ashamed.'

The incorrect addition of de can happen between the verb and the object in a sentence, as in sentence (6), or at the end of a clause, as in sentence (7). Such diversity can be ascribed to the sophisticated functions of the particle de: de can be used as an aspectual particle in sentences, but its usage is not quite the same as inflectional affixes in those richly inflected languages, which results in misuse. The usage of de as a modal particle is also very frequent, and it is easy for learners to add it to the end of a sentence when unnecessary, thus causing redundancy.

3.3. Misordering

Misordering errors happen when a morpheme or group of morphemes are not placed correctly in an utterance [1]. The misordering errors in *de* structures mainly occur in the position of the attribute and adverbial.

(8)这　　都是　　我　　错的。
zhe　　dou-shi　wo　　cuo-de
this　　all-is　　I　　fault-NMLZ
'This is all my fault.'

(9)常常　　天空　　是　　黄色的。
changchang　tiankong　shi　huangse-de
Usually　　sky　　is　　yellow-NMLZ
'The sky is usually yellow.'

In the contemporary Chinese language, the attribute and adverbial must precede the head, and an optional de should be placed between them as a marker of modification. Sentence (8) reverses the order of the noun 错(cuo, 'fault') and the marker de. The error in sentence (9) is related to the adverbial. The adverbial is usually located between the subject and the predicate in Chinese, and only a small number of adverbials that indicate time, place, or purpose can be placed at the beginning of the sentence, and the adverb 常常(changchang, 'always') is not one of them. Therefore, the correct order of sentence (9) should be 天空常常是黄色的, reversing the order of the subject and the adverbial.

3.4. Misselection

Misselection is the use of the wrong form of a structure or morpheme [1]. Similar pronunciations, meanings or syntactic features (patterns of characters should also be considered when analyzing Chinese) of different linguistic units lead to misjudgements, for example, in the choice of tense, aspect, word class, voice or lexis [1,10]. Misselection errors of the de structure can be roughly divided into two subtypes: misselection among three homophonic de and misselection among the de structure and other structures.

3.4.1. Misselection among Three Homophonic De

(10)所以　不要　　拼命的　　　找　　好　　丈夫。
suoyi　bu-yao　pinming-de　　zhao　hao　zhangfu
so　　not-do　desperately-NM　find　good　husband
'So do not search for a good husband desperately.'

(11)过的好吗?
guo-de-hao-ma
live-NM-good-AUX.Q
'How are you doing?'

As mentioned in section 1.1, de works as the marker of modification in subordinative phrases. However, such a statement is not precise enough because in the contemporary Chinese language, there are three homophonic de among which clear-cut boundaries exist:的(de) is used in nominal phrases while 地(de) and 得(de) work as markers of modification in adverbial-verb structures and

verb-complement structures respectively. Due to their contrastive distribution patterns, three different de can be viewed as conditional variants of the same entity, and this is exactly where misselection errors are prone to happen. The same pronunciations and highly close syntactic features cause great confusion to learners and result in the misuse of three de. Both sentence (10) and sentence (11) substitute the correct forms with the noun marker 的(de).

3.4.2. Misselection among the De Structure and Other Structures

(12)我	在	宿舍	做	很多	活动,	有的	学习	汉语，	有的	看	电视。
wo	zai	sushe	zuo	hen-duo	huodong	you-de	xuexi	hanyu	you-de	kan	dianshi
I	in	dormitory	do	very-many	activity	have-NMLZ	study	Chinese	have-NMLZ	what	TV

'I did lots of things in the dormitory, sometimes studying Chinese, sometimes watching TV.'

In Chinese, the structure 有的…有的… is used to enumerate people or things of homogenous types, generally equivalent to the structure some…some… in English. However, based on the context of sentence (12), subject 我(wo, 'I') is the agent of all these actions, so what the writer had intended to convey is 有时候学习汉语，有时候看电视(sometimes studying Chinese, sometimes watching TV), with the adverbial 有时候(youshihou, 'sometimes') preceding verb-object phrases.

3.5. Blend

Blend refers to the situation when two alternative grammatical forms are combined to produce an ungrammatical one [2]. The learner is irresolute between two possible options and ends up blending them together instead of choosing either.

(13)比如	你	一定	努力	学习、	好好	看书、	每天	要	复习	什么	等。
biru	ni	yiding	Nuli	xuexi,	haohao	kanshu,	meitian	yao	Fuxi	shenme	deng
such.as	you	must	Hard	study,	well	read,	everyday	need	Review	what	etc.

'For example, you must study hard, read earnestly, review everyday and so on.'

The error happens in the rightmost part of the sentence as the writer combined two words 什么的 (shenmede, 'and so on')and 等 (deng, 'etc.') together to coin a non-existent word 什么等 (shenmedeng). The compound 什么的 is constructed by the combination of the interrogative pronoun 什么 (shenme, 'what') and de, used after a list to show that there are other things that one could have mentioned. Both 什么的 and 等 can convey the meaning of enumeration in Chinese, but the simultaneous use of them will cause redundancy and thus is unacceptable. Blends demonstrate learners' unfamiliarity with different structures, especially those similar ones.

4. The Causes of Errors in De Structures

4.1. L1 Transfer

There is general agreement that previous knowledge of L1 will affect the process of second language acquisition, which was referred to as L1 transfer by many scholars. L1 transfer can facilitate learning when there is some degree of similarity between linguistic units in L2 and equivalent units in L1 [2]. However, L1 transfer can also interfere with second language acquisition, and this section will concentrate on the negative effect of L1 transfer. Based on the dichotomy proposed by Hammerly,

negative transfer of L1 appears in two situations [11]; if the learner's attempt to transfer results in inappropriate behavior in the new context, intrusive interference happens [11]. A typical example comes from L1 Korean learners: In Chinese, when a verb act as the X component in a de structure, the whole structure generally refers to the agent, patient or instrument of the central verb, but not the action conveyed by it. However, in Korean, the 것(ged) structure, which is similar in function to the de structure, can refer to the action itself in the same situation [12]. Therefore, L1 Korean learners always insert the reductant de after verbs to emphasize them, which is ungrammatical in Chinese. Another term given by Hammerly is inhibitive interference, occurring when there is no equivalent in learners' L1 [11]. Take L1 English learners, for example: Whether as a nominalizer or a modification of marker, the particle de is completely new to L1 English learners, and so is the de structure; learners have no alternative recourse but to omit them in their output of L2. This somehow explains the occurrence of omission errors in section 2.1.1; learners chose to avoid the use of unfamiliar de but still yield the errors, as de structures are the only acceptable constructions here.

4.2. Overgeneralization

Language consists of an infinite set of sentences generated by a finite set of rules, and the learners must make correct generalizations based on the finite sample given in the input [13]. However, the partial regularities in language always lead to overgeneralization: the improper generalization of one particular rule to its exceptions [13]. Many types of errors in de structures have a lot to do with overgeneralization. In Chinese, a de can be attached after verbs to express a completed action, as in the sentence (14).

(14)我们 乘 飞机 去的 北京。
women cheng feiji qu-de beijing
we ride plane go-AUX.PRF Beijing
'We have been to Beijing by plane.'

In this sense, de works as an aspectual particle that follows verbs to form perfect aspects. However, this usage of de is highly limited and only appears in some particular situations. CSL learners sometimes overlook those limitations and overgeneralize the uses of aspectual de.

(15)我 第一次 学习汉语的 是 六岁的时候。
Wo diyi-ci xuexi-hanyu-de shi liu-sui-de-shihou
I First study-Chinese- AUX.PRF is six-year-NM-time
'I first learned Chinese when I was six years old.'

The context induces the addition of aspectual de, which is completely reductant here.

5. Conclusion

This article demonstrates the processes and findings of the investigation in errors yielded by CSL learners when using de structures. By random sampling and further sifting, 155 sentences are selected as valid samples of the de structure, among which 87 samples are incorrect in the use of de structures, with an error rate of 56%, and the three most problematic forms are subordinative phrase (18%), verb (18%) and adverbial (16%). Based on the distortion of the intended correct form, errors in de structures can be categorized into five main types and five subtypes: omission (omission of the de, omission of the head, omission of other constituents), overinclusion, misordering, misselection (misselection among three homophonic de, misselection among the de structure and other structures) and blend. The contrastive analysis among Chinese, Korean and English illustrates how L1 transfer

interferes with second language acquisition, and the overgeneralization of aspectual de serves as an example of intralingual errors, on the other hand. The error analysis based on a large-scale interlanguage corpus makes it possible for CSL teachers and textbook designers to re-examine the current teaching materials and pedagogies in a more empirical way, which can be helpful in the teaching of Chinese.

References

[1]   Dulay, H., Burt, M. and Krashen, S.D. (1982). Language Two. New York: Oxford University Press.
[2]   James, C. (2013). Errors in Language Learning and Use: Exploring Error Analysis. New York: Routledge.
[3]   Zhu, D. X. (1961). On De. [Shuo De]. Studies of the Chinese Language, 12, 34-37.
[4]   Chao, Y. R. (2011). A Grammar of Spoken Chinese. Beijing: The Commercial Press.
[5]   Huang, B. R., & Liao, X. D. (2011). Contemporary Chinese Language. [Xiandai Hanyu]. Beijing: Higher Education Press.
[6]   Zhang, B. L., Cui, X. L. (2022). Features and Functions of Global Chinese Interlanguage Corpus. Chinese Teaching in the World, 1, 90-100.
[7]   Liu, K. Z. (2015). Multi-angle Analysis of the De-Phrases on the Base of Corpus. Guangxi Normal University.
[8]   Chen, Y. J., Sousa, H., Wang, J., Ni, X. X., Li, X. P., Chen, W. R., Chappell, H. (2014) Leipzig Glossing Rules and Its Application in the Annotations of Chinese Grammar. Dialect, 1, 1-13.
[9]   Lü, S. X. (1980). Eight Hundred Words in Modern Chinese. [Xiandai Hanyu Babai Ci]. Beijing: The Commercial Press.
[10] Chuang, F. Y., & Nesi, H. (2006). An analysis of formal errors in a corpus of L2 English produced by Chinese students. Corpora, 2, 251-271.
[11] Hammerly, H. (1991). Fluency and Accuracy. Bristol: The Longdunn Press.
[12] Wang, H. (2016). The Comparison of De Structure of Chinese and Ged structure of Korean. Eastern Forum, 4, 112-116.
[13] Yip, V. (1995). Interlanguage and Learnability: From Chinese to English. Philadelphia: John Benjamins Publishing Company.