RETRIEVING LINGUISTIC INFORMATION FROM A CORPUS ON THE EXAMPLE OF NEGATION IN CHINESE

Ľuboš GAJDOŠ
Comenius University in Bratislava, Slovakia
lubos.gajdos@uniba.sk

Abstract

The paper deals with corpus analysis of negation in Chinese, namely the negatives bù 不 and méi/méiyǒu 没/没有. The adverbs BU and MEI are two of the most frequent negatives in Chinese. The aim of this study is to present statistical data together with linguistic analysis. The results provide empirical evidence of discrepancy between "authentic" language data versus linguistic prescription with practical implications for second-language acquisition. The findings inter alia suggest a new approach to verb categorisation.

Keywords: Chinese language; corpus linguistics; quantitative description; negation; potential complements

Povzetek

Članek obravnava korpusno analizo negacije v kitajščini, pri čemer se avtor osredotoča na prislova bù 不 in méi/méiyǒu 没/没有, ki sta najpogostejši nikalnici v sodobnem kitajskem jeziku. Namen prispevka je predstaviti statistične podatke v povezavi z jezikoslovno analizo. Rezultati študije prinašajo empirične dokaze o neskladju med jezikovno rabo in jezikovnimi normami, ta spoznanja pa je moč uporabiti tudi pri poučevanju kitajščine kot tujega jezika in za premislek o drugačnem pristopu na kategorizacijo glagolov.

Ključne besede: kitajščina; korpusno jezikoslovje; kvantitativni opis; negacija; zmožnostna dopolnila

Acta Linguistica Asiatica, 9(2), 2019.
ISSN: 2232-3317, http://revije.ff.uni-lj.si/ala/
DOI: 10.4312/ala.9.2.103-115

1 Introduction

Generally speaking, there are a number of negatives in modern Chinese.[1] In this article only two negative adverbs, namely *bù* 不 and *méi/ méiyǒu* 没/没有[2] (hereafter referred to as BU and MEI), are discussed. The Hanku corpus is used[3] as the primary source of language material and statistical data. As the intention is to mainly use the corpus-driven[4] approach to studying of negation, thus the previous linguistics research on this topic is left aside.

Let us start with some basic queries:[5]

[tag="AD" & word="不"]
[tag="VV|AD" & word="没|没有"][6]

The results are 7371142 (9897.85 per million),[7] 686352 (921.62 per million)[8] respectively. These numbers only tell that the token BU is approximately 10 times more frequent than MEI. The difference is even more pronounced when searching in a certain variety of Chinese, e.g. in the corpus of legal Chinese, the occurrence of BU is 45254 (6 281.56 per million), the occurrence of MEI 720 (99.94 per million). Let us take a closer look at the tokens that collocate with these negatives. The following queries should return collocates at the position 1 on the right side:[9]

[tag="AD" & word="不"][]
[tag="AD"& word="没|没有"][]

[1] For details see e.g. Liu (2004, pp. 253–258).

[2] For the sake of simplicity, both negatives *mei* and *meiyou* are treated as two forms of one negative, namely MEI. On the other hand, their collocative partners may differ because of e.g. prosodic factors.

[3] See more in Gajdoš, Garabík and Benická (2016, pp. 53–65).

[4] See more in Baker, Hardie and McEnery (2006, p. 49).

[5] In this article, the Corpus Query Language (hereafter CQL) is used to search for collocations. With CQL, complex criteria can be set to find one or many tokens. Criteria for each token must be between a pair of square brackets [ ], e.g. [attribute="value"]. See more at https://www.sketchengine.eu/documentation/cql-basics/

[6] As there are more tags (e.g. VV = verbs, VE = YOU as the main verb, AD = adverbs) dedicated to tokens *mei* 没 and *meiyou* 没有, it is rather difficult to accurately determine the value of the negative MEI. Thus, I only use tags VV, AD in this article. For more details on the tagset see Fei (2000, pp. 4–35).

[7] Unless stated otherwise, frequencies are given in absolute occurrence in the Hanku corpus.

[8] The occurrence of *mei* 没 is 342190 (459.49 per million) and 344162 (462.14 per million) for *meiyou* 没有.

[9] The regular expressions may match the following patterns:

adverb *bu* + any token or
adverb *mei/meyou* + any token.

As it is rather difficult to identify the collocates at the position further to the right using only POS tags, this topic will be left for future research. See e.g. Gajdoš (2018).

The results are summarised in the tables below.[10]

Table 1: The most frequent POS at the position 1 (Corpus: web-zh)

# Query: word, [tag="AD" & word="不"][] # Query: word, [tag="AD|VV" & word="没|没有"][]			
不		没|没有	
No. of results: 7371142		No. of results: 686352	
tag	frequency	tag	frequency
AD VV	5092363	**VV VV**	326984
AD VA	827140	**AD VV**	193781
AD VC	695511	**VV AD**	35762
AD AD	444055	**VV P**	30138
AD P	155025	**AD AD**	16246
AD PU	25581	**VV AS**	15136
AD AS	19730	**VV PU**	9902
AD JJ	19525	VV CD	5771
AD PN	13121	VV PN	5350
AD NN	10044	VV VA	4975
AD BA	9276	VV BA	4870
AD VE	8798	VV NN	4449
AD LB	8553	VV DT	4223
AD CD	7535	AD VA	3511
AD SB	6815	AD P	3488
AD NR	5930	VV SB	2795
AD DT	4226	VV DEC	2660
AD DEC	3170	VV LB	1914
AD DEV	2380	AD SB	1541
AD SP	1987	VV DER	1460
AD LC	1744	AD BA	1334
AD M	1709	VV JJ	1209
AD MSP	1545	VV NR	1106
AD NT	1097	VV NT	959
AD OD	1032	AD CD	861
AD CC	1023	AD PN	691
AD CS	1021	AD VC	663
AD DEG	941	VV VC	526
AD DER	118	VV OD	499
AD ETC	88	VV LC	380
AD FW	45	AD VE	354
AD IJ	14	AD LB	338

[10] The results are calculated using the NoSketch Engine UI – Node tags.

The table indicates different collocability for the negative BU and MEI, e.g. the negative BU exhibits a strong preference for copulas (here VC).[11] For practical reasons, only the POS tags, which are more frequent than 1% of each group (here in bold), are included in the analysis. The PU tag is also to be excluded from further analysis as it stands for punctuation.

Table 2 shows 10 of the most frequent collocates for each negative. The results are calculated using the NoSketch Engine UI – Node forms.[12]

Table 2: The most frequent tokens at the position 1 (# Corpus: web-zh)

# Query: word, (meet [tag="VV|VA|VC|AD|P"]2:[tag="AD" & word="不"]-1 -1)
# Query: word, (meet [tag="VV|AD|P|AS"]2:[tag="AD|VV" & word="没|没有"]-1 -1)

不		没|没有	
No. of results: 7214094		No. of results: 621555	
word	frequency	word	frequency
是	693223	想到	56260
能	377529	看	14539
会	311133	**能**	13999
知道	288756	了	12223
要	227692	见	9843
存在	125328	在	9249
可	125212	说	9215
到	124252	想	9141
得	91523	看到	7189
敢	85127	用	6744

At first sight, it is surprising that the collocation MEI+ *néng* 能 is the third most frequent, despite the fact that most grammars and textbooks deny this possibility.[13] Similar findings may provide the impetus for further research which would take greater

[11] The co-occurrence of MEI+VC is caused by the misspelling of the character *shi* in most cases, e.g. *mei shi* 没是 instead of *mei shi* 没事.

[12] To find collocative partners of both negatives, the operator *meet* is used. That means that the corpus is search for the following patterns:

adverb BU + verbs (VV) or
adverb BU + adjectives (VA) or
adverb BU + copulas (VC) or
adverb BU + adverbs (AD) or
adverb BU + prepositions (P).

See more at https://www.sketchengine.eu/documentation/cql-meet-union/

[13] There are some exceptions, e.g. Švarný and Uher (2014, p. 48) describe this phenomenon and Liu (2004, p. 257) also suggests this possibility, however, they do not further elaborate this point.

account of actual language use. The “new” grammars or textbooks should be then based on such research.

After searching the first hundred examples manually, it turns out that the co-occurrence of some tokens with BU is higher than one would expect based on the frequency of affirmation, e.g. *liǎo* 了 (70691), *zhù* 住 (65400), *qǐ* 起 (40086) etc., furthermore, these verbs typically serve as so-called complements.[14] That means that only tentative conclusion may be drawn from this evidence, nevertheless, it should play a role when comparing the overall frequency of both negatives. I discuss this topic further in the chapter *Potential complements*.

2 Potential complements

Let us return to the examples that have been mentioned in Chapter 1 and analyse them.

(1)	他	**办**	**不/AD**	**了/VV**	此事
	Tā	bàn	bù/AD	liǎo/VV	cǐ shì
	He	manage	BU-neg.	compl.	this matter

‘He cannot do this.’

(2)	杰克终于	**忍**	**不/AD**	**住/VV**	说了
	jiékè zhōngyú	rěn	bù/AD	zhù/VV	shuō le
	Jack finally	endure	BU-neg.	compl.	speak LE

‘Jack finally couldn't help saying it.’

(3)	这辈子房子	**买**	**不/AD**	**起/VV**	了
	zhè bèizi fángzi	mǎi	bù/AD	qǐ/VV	le
	this life house	buy	BU-neg.	compl.	LE

‘(One) cannot afford to buy a house for the entire life.’

(4)	对于双方不	**能**	**不/AD**	**说/VV**
	duìyú shuāngfāng bù	néng	bù/AD	shuō/ VV
	for to both sides not	able to	BU-neg.	speak

‘regarding (things that) both sides cannot but speak’

Randomly selected samples suggest that many examples may be considered as so-called potential complements with the “morphological” structure VV + BU + VV while

[14] See e.g. Yip (2009, pp. 234–241).

the first morpheme (verb) is not equal to the third. The following query meets this condition:[15]

```
(meet (meet 1:[tag="VV"][tag="AD" & word="不"]-1 -1) 2:[tag="VV|VA"]-2 -2) &
1.word!=2.word
```

The examples below show that the regular expression does not always match the desired pattern and therefore must be modified.

(5) 那个人呢叫李一， **知** **不** **知道/VV**？
Nàgè rén ne, jiào Lǐ Yī, zhī bù zhīdào
That person is Li Yi know BU-neg. know
‘Do you know that that person is called Li Yi?’

(6) 您 **能** **不** **能够/VV** 再具体地跟我们讲一下？
Nín néng bù nénggòu zài jùtǐ de gēn wǒmen jiǎng yīxià?
You able to BU-neg. able to tell us more specifically
‘Can you tell us this again more specifically?’

(7) 一定程度上，不 **能** **不** **说/VV**
Yīdìng chéngdù shàng, bù néng bù shuō/VV
to a certain extent, not able BU-neg. speak
‘To a certain extent, one cannot but speak.’

(8) 一个 说 不 **要/VV**
Yīgè shuō bù yào
one say BU-neg. want
‘One says *no*.’

(9) 这些人 找 不 **到/VV** 工作
Zhèxiē rén zhǎo bù dào/VV gōngzuò
these people find BU-neg. compl. work
‘These people cannot find work.’

There is a point worth noting here as well – auxiliary verbs (e.g. modal verbs) must be removed from the search pattern. As there is no dedicated tag for modal or auxiliary

[15] This regular expression matches the following pattern:
verb2/adjective2 + BU + **verb1** and verb1 ≠ verb2, the verb1 is KWIC (Key Word in Context).

verbs (except VE, VC), each of the verbs must be enumerated in the query with the attribute "word".[16] A double negative must be excluded too. The refined query is:[17]

(meet (meet (meet 1:[tag="VV"& word!= "要|能" "& word="(?i).{1,2}"]
[tag="AD" & word="不"]-1 -1) 2:[tag="VV|VA"& word!= "要|能"]-2 -
2)[word!="不"]-3 -3) & 1.word!=2.word

The following table shows the result. The overall frequency is 828224 (1112.13 per million).

Table 3: The most frequent potential complements – the negative form (# Corpus: web-zh)

# Query: word,(meet (meet (meet 1:[tag="VV"& word!= "会|能|应该|该|必须|可以|可|应当|可以|应|能|能够|必须|须|要|可能|会|需要|愿意|敢|该|需|知道" & word="(?i).{1,2}"][tag="AD" & word="不"]-1 -1) 2:[tag="VV|VA"& word!= "会|能|应该|该|必须|可以|可|应当|可以|应|能|能够|必须|须|要|可能|会|需要|愿意|敢|该|需|知道"]-2 -2)[word!="不"]-3 -3) & 1.word!=2.word

word	Frequency
到	111457
了 *liǎo*	65697
住	59793
起	36774
得 *dé*	33857
出	31377
上	29568
开	22326
过	19412
见	12632
着 *zháo*	10366
出来	10114
懂	9877

[16] The query above contains only two of these verbs, the others are present here, e.g. 能|应该|该|必须|可以|可|应当|应|能够|必须|须|要|可能|会|可|需要|愿意|敢|该|需 etc. The limit for the length of the tokens is set to 1 or 2 by the expression: "word="(?i).{1,2}".

[17] The regular expression means that the corpus is searched for the following pattern:
token (not BU) + verb2 (not 要 nor 能) + adverb BU + mono- or disyllabic **verb1** (which is not 要 nor 能) and verb1 ≠ verb2. Only the verb1 is KWIC in the concordance and other tokens are used as contextual filters. See more at https://www.sketchengine.eu/documentation/cql-meet-union/

The result of the affirmative form might be achieved by the same query with only minor modification:[18]

```
(meet (meet (meet 1:[tag="VV"& word!= "要|能" "& word="(?i).{1,2}"]
[tag="DER"]-1 -1) 2:[tag="VV|VA"& word!= "要|能"]-2 -2)[word!="不"]-3 -3)
& 1.word!=2.word
```

The total frequency of 167822 (225.35 per million) clearly shows that the occurrence of the affirmative form is far less frequent. This fact only validates the previous assumption mentioned in the literature.[19] The following list contains a sample of the most frequent verbs: 上 11598, 起 10084, 到 9769, 住 7 614, 出 7607, 出来, 3736, 见 2977 etc.

If we move back to the calculation of the overall frequency of BU, the value of the negative form of potential complements (1112.13 per million) should be subtracted from the total frequency, i.e. 8785.72 per million. Needless to say, these are only approximate numbers and further research is required.

3 Verb collocates

The first chapter discusses the collocability of the negative BU and MEI. In this chapter, I further explore this topic. When comparing the total frequency of BU vs. MEI, some considerations should be taken into account, i.e. some verbs/adjectives collocate with BU only, some registers use only a limited number of MEI etc.

After saving the results as a text file (from the NoSketchEngine UI), I proceed to test the 2 lists[20] for the duplication[21] and calculate the average value of co-occurrence. When comparing two lists for duplication in the spreadsheet program, there are many tokens in the MEI list which are marked as they have no counterpart in the BU list. This might cause surprise at first since one would expect only tokens from the BU list not having a counterpart. The explanation is rather simple: (1) most of these tokens have a disyllabic morphological structure (V+X), e.g. 找到, 看到 and cannot be paired with their monosyllabic counterpart in the BU list by the spreadsheet program (e.g. 找, 看) or (2) the frequency of the BU counterpart is below the lowest frequency of samples (see footnote 13).

[18] There is a dedicated tag for the 得 *de*-marker, i.e. DER.

[19] See e.g. Liu (2004, p. 583).

[20] Each list contains the 1000 most frequent verbs that collocate with BU and MEI.

[21] This might be done in MS Excel, LibreOffice Calc or any spreadsheet program.

Table 4: The 10 most frequent verbs collocating with BU and MEI (# Corpus: web-zh)

<table>
<tr><td colspan="4"># Query: word,(meet [tag="VV"]2:[tag="AD" & word="不"]-1 -1)
Query: word,(meet 1:[tag="VV" & word="(?i).{1,2}"][tag="AD|VV" & word="没|没有"]-1 -1)</td></tr>
<tr><td colspan="2">不</td><td colspan="2">没|没有</td></tr>
<tr><td>word</td><td>Frequency</td><td>word</td><td>Frequency</td></tr>
<tr><td>能</td><td>377442</td><td>想到</td><td>56260</td></tr>
<tr><td>会</td><td>311129</td><td>看</td><td>14539</td></tr>
<tr><td>知道</td><td>288756</td><td>能</td><td>13999</td></tr>
<tr><td>要</td><td>227691</td><td>见</td><td>9843</td></tr>
<tr><td>存在</td><td>125328</td><td>说</td><td>9215</td></tr>
<tr><td>可</td><td>124875</td><td>想</td><td>9141</td></tr>
<tr><td>到</td><td>122930</td><td>看到</td><td>7189</td></tr>
<tr><td>得</td><td>91458</td><td>用</td><td>6086</td></tr>
<tr><td>敢</td><td>85108</td><td>去</td><td>5602</td></tr>
<tr><td>知</td><td>78854</td><td>来</td><td>5371</td></tr>
</table>

The results indicate that:

- From the list of the 1000 most frequent tokens (verbs) with the negative BU, 619 tokens collocate with MEI too, yet from the 100 most frequent tokens, there are 69 of them; the rest are e.g. the following tokens: 知, 行, 愿, 愿意, 肯, 应, 信 etc. that co-occur with BU only;
- From the list of the 100 most frequent tokens (verbs) with the negative MEI, a few preferably collocate with MEI, e.g. 发现, 料到, 必要, 开始, etc.;[22]
- The lower the frequency of a token in the BU list, the less frequent it collocates with both negatives;
- Generally speaking, the co-occurrence of the negative MEI with the same verb is about 2.5-time less frequent as with the BU negative, however, statistical data reveals great disparities between tokens (see table 5). That is to say that verbs on the left side of the table collocate almost always with the negative BU, on the other hand, verbs on the right side almost exclusively collocate with the negative MEI.

[22] This may be seen from the following comparison: the query [word="没|没有" & tag="VV|AD"][word="发现"] with the frequency of 4542 (6.10 per million) and the query [word="不" & tag="AD"][word="发现"] 62 (0.08 per million).

Table 5: Collocability of verbs (# Corpus: web-zh)

Preference for BU		Preference for MEI	
word	**ratio**	**word**	**ratio**
知道	1511,8	想到	0,005
存在	858,4	放松	0,182
会	781,7	看到	0,201
住	408,7	留下	0,277
可	325,2	出现	0,290
在	307,7	进入	0,314
起	301,4	选择	0,323
合	278,5	感觉	0,340
应该	236,3	受到	0,342
了	235,4	表现	0,385

4 Adjective and adverbs collocates

This chapter focuses on the collocability of adjectives and adverbs and the same searching methods are used.

As for the adjectives, a brief look at the given statistical data (827140 or 1110.67 per million vs. 8486 or 11.39 per million; see table 6) demonstrates that adjectives (almost) always collocate with the negative BU. The exceptions here may be considered as phrases.

Table 6: Collocability of adjectives (# Corpus: web-zh)

<table>
<tr><td colspan="4"># Query: word,(meet [tag="VA"]2:[tag="AD" & word="不"]-1 -1)
Query: word,(meet [tag="VA"]2:[tag="AD|VV" & word="没|没有"]-1 -1)</td></tr>
<tr><td colspan="2">不
No. of results: 827140</td><td colspan="2">没|没有
No. of results: 8486</td></tr>
<tr><td>word</td><td>Frequency</td><td>word</td><td>Frequency</td></tr>
<tr><td>好</td><td>71999</td><td>错</td><td>1978</td></tr>
<tr><td>美</td><td>58946</td><td>成功</td><td>884</td></tr>
<tr><td>多</td><td>46448</td><td>好气</td><td>812</td></tr>
<tr><td>够</td><td>34917</td><td>多</td><td>374</td></tr>
<tr><td>一样</td><td>28451</td><td>真正</td><td>329</td></tr>
<tr><td>大</td><td>23685</td><td>必要</td><td>257</td></tr>
<tr><td>美观</td><td>18118</td><td>好</td><td>251</td></tr>
<tr><td>高</td><td>14844</td><td>明确</td><td>205</td></tr>
<tr><td>容易</td><td>14831</td><td>成熟</td><td>195</td></tr>
<tr><td>小</td><td>13191</td><td>好好</td><td>180</td></tr>
</table>

The situation with regard to adverbs is a little different. While the results indicate a strong tendency to the negative BU, yet both negatives may be used.

Table 7: Collocability of adverbs (# Corpus: web-zh)

# Query: word,(meet [tag="AD"]2:[tag="AD" & word="不"]-1 -1)
# Query: word,(meet [tag="AD"]2:[tag="AD|VV" & word="没|没有"]-1 -1)

不		没|没有	
No. of results: 444055		No. of results: 52008	
word	frequency	word	frequency
再	76724	那么	6480
太	39336	再	4637
一定	22521	这么	4062
就	19298	完全	2989
只	17114	多	2855
曾	13088	怎么	2316
单	12286	真正	1889
正	10410	不	1541
多	8623	太	988
怎么	8621	甚么	920

5 Conclusion

To begin with, statistical data given in this study should only be taken as exhibiting a general tendency and not as a fully accurate description of “real” language. It should also be pointed out that this paper only examines the occurrence of negatives at the first position to the left of collocates. In this respect, new methods should be devised for solving issues addressed here, e.g. the problem with the POS annotation and its error rate which may significantly affect statistical data or the problem with identifying the difference between the negative MEI and the verb *yǒu* 有 (with the tag VE) etc. This leads us to the questions how to interpret the results in light of these points and what valuable results this study brings.

Firstly, when comparing results of both negatives, it seems that some verbs described as “auxiliary” or “modal” tend to collocate with the negative MEI more often than stated by language prescription. On the other hand, empirical data support the claim that adjectives only collocate with the negative BU. As for adverbs, there is still a strong preference for BU, but because I do not consider adverbs as a “true” collocate to negatives (rather as part of a bigger structure), this question should be explored in future research.

Let us now move on to the negative MEI. There are many verbs that preferably collocate with MEI rather than with BU. A closer look at the results reveals that their morphological structure is disyllabic and the left morpheme is often a so-called "resultative complement" (*jiéguŏ bŭyŭ* 结果补语). This finding may imply that the category of verbal aspect and tense[23] deserves closer attention. That means if MEI is regarded as past time marker, these verbs are commonly used in past tense and the present tense (with BU) may describe the situation as a condition or future tense. A similar phenomenon is also observed in some Slavic languages, where the present and preterite of perfective verbs fulfil these functions too (e.g. compare the present perfective form "urobím" vs. the past perfective form "urobil" in Slovak). This suggests that these verbs in Chinese might be treated as *perfective*. In order to fully explore this topic, the marker *le* 了, as a counterpart to the negative MEI, should be included in an comparative analysis. There is a very detailed, corpus-based study conducted on this subject by Petrovčič (2009), *Operator Le in Chinese* worth noting here.

To conclude, the article shows how to use a corpus when searching for evidence of some language phenomena. As for negation in Chinese, the paper only suggests a different approach to this subject and additional research is needed.

References

Baker P., Hardie A., & McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Fei, X. (2000). *The Part-Of-Speech Tagging Guidelines for the Penn Chinese Treebank (3.0)*. Available at https://www.cs.brandeis.edu/~clp/ctb/posguide.3rd.ch.pdf

Gajdoš, Ľ., Garabík, R., & Benická, J. (2016). The New Chinese Webcorpus Hanku – Origin, Parameters, Usage. *Studia Orientalia Slovaca, 15*(1), 53–65.

Gajdoš, Ľ. (2018). Korpusová analýza adverbiále v právnej čínštine [Corpus Analysis of adverbial complements in legal Chinese]. In T. Guldanová (Ed.), *Kontexty súdneho prekladu VII* (pp. 27–39). Bratislava: Univerzita Komenského.

Liu, Y. [刘月华] et al. (2004). *Practical Chinese Grammar* [实用汉语语法]. Beijing: Shangwu yishuguan.

Petrovčič, M. (2009). *Operator Le in Chinese.* Saarbrücken: VDM Verlag.

Petrovčič, M. (2017). Traditional and Contemporary Approaches to Chinese Particles. In S. Bračič & M. Petrovčič (Eds.), *Partikeln überall: Deutsch - Slowenisch - Chinesisch*, (pp. 103-122). Ljubljana: Znanstvena založba Filozofske fakultete.

Sketch Engine. Available at https://www.sketchengine.eu

Švarný, O., & Uher, D. (2014). *Prozodická gramatika čínštiny* [Prosodic Grammar of Chinese Language]. Olomouc: Univerzita Palackého.

Yip, P., & Rimmington, D. (2009). *Basic Chinese.* Abingdon: Routledge.

[23] See also Petrovčič (2017, pp. 108–109).

Appendix: The Hanku tagset[24]

Tag	English	Example
AD	adverb	也
AS	aspect particle	着
BA	preposition BA in ba-construction	把
CC	coordinating conjunction	和
CD	cardinal number	十五
CS	subordinating conjunction	如果
DEC	markers – nominalizer	吃的
DEG	genitive marker	他的
DER	resultative DE 得	说得
DEV	manner DE 地	公正地
DT	determiner	这
ETC	et cetera	等
FW	foreign word	ISBN
IJ	interjection	喂
JJ	other noun-modifier	女
LB	preposition BEI in long bei-construction	被
LC	localizer	上
M	measure word	个
MSP	other particle	所
NN	noun	记者
NR	proper noun	英语
NT	temporal noun	今年
OD	ordinal number	第三
ON	onomatopoeia	哈哈
P	preposition	从
PN	pronoun	我
PU	punctuation	。
SB	preposition BEI in short bei-construction	被
SP	sentence-final particle	了
VA	predicative adjective	大
VC	copula	是
VE	verb 有/没有/无 as the main verb	有
VV	verb	说

[24] For details see Fei (2000, pp. 4–35).