The *de*-construction in Wenzhounese: how it differs from the *bǎ*-construction in Mandarin

Chen Xie[1]

Abstract

The canonical word order in Mandarin is SVO, but it becomes "S *bǎ* O V" in the *bǎ*-construction, where *bǎ* vaguely expresses the affectedness of the object. How to analyze the *bǎ*-construction is one of the most controversial topics in Chinese grammar. This paper approaches the issue from a fresh perspective by presenting new data from Wenzhounese, where the *de*-construction is parallel to the *bǎ*-construction. Although these constructions share a considerable degree of similarity, they differ in the following key respects. First, they differ in the strategies used to topicalize the post-*de/bǎ* noun phrase (NP). Second, the quantifier taking scope over the post-*de/bǎ* NP has distinct distributions. Third, Wenzhounese has a double-*de* construction unattested in Mandarin. Based on these observations, I offer a formal analysis of the *de*-construction within the framework of Lexical Functional Grammar. This encompasses the lexical entries, constituent-structure, and functional-structure for *de*, which will capture the differences between *de* and *bǎ* and shed new light on the *bǎ*-construction in Mandarin.

Keywords Wenzhounese · Mandarin · Syntax · The *bǎ*-construction · LFG

1 Introduction

Wenzhounese is a subgroup of the Wu dialect (Zhengzhang and Zheng 2015), mainly spoken in Wenzhou City, Zhejiang Province, China. Its canonical word order is SVO, but in the presence of an ostensible object marker *de*, the relative order of V and O is reversed, as shown in (1).[1] The (un)acceptability of the Wenzhounese sentences in

---

[1] Mandarin data are marked with "⟨M⟩" and transcribed with *pinyin*. Wenzhounese data are labelled with "⟨W⟩" and transcribed with the IPA, based on a variety spoken in Rui'an, Wenzhou. Tones are represented by Chao's digits (Chao 1930) and the tone values are largely based on Xu (2019). Unless otherwise specified, only surface tones are provided in the examples, while underlying tones are omitted (see Sect. 2.6 for tone sandhi). Sentences labelled "⟨WSC⟩" are drawn from the Wenzhou Spoken Corpus (Lin and Newman 2004), with minor revisions when appropriate.

✉ Chen Xie
 chen.xie@ling-phil.ox.ac.uk

1 Faculty of Linguistics, Philology and Phonetics, University of Oxford, Oxford, UK

this paper has been checked with four native speakers of Wenzhounese (excluding the author, who is also a native speaker of Wenzhounese).

(1) a. $l\varepsilon^0$-$t\varepsilon\varepsilon^{44}$ $tsei^0$ $sou^{53}t\varepsilon i^{44}$ $k^h\mathfrak{z}^0$ $l\varepsilon^0$-fio^{31} $\langle$W$\rangle$
       old-Zhang lend phone   to   old-Huang
       'Old Zhang lent his phone to old Huang.'
     b. $l\varepsilon^0$-$t\varepsilon\varepsilon^{44}$ de^0 $sou^{53}t\varepsilon i^{44}$ $tsei^{53}$ $k^h\mathfrak{z}^0$ $l\varepsilon^0$-fio^{31} $\langle$W$\rangle$
       old-Zhang DE phone   lend to   old-Huang
       'Old Zhang lent his phone to old Huang.'

(1a) exemplifies the VO order, where the verb *tsei* 'lend' precedes the object *soutɕi* 'phone'. In (1b), *soutɕi* precedes *tsei*, resulting in the word order *de* + OV. Linearly, the *de*-construction takes the form "Subject *de* Object Verb".

Essentially, *de* is the counterpart of Mandarin *bǎ*, a morpheme arguably expressing the affectedness of the object (cf. Li 2006) and resulting in the linear order "Subject *bǎ* Object Verb" similar to (1b). Additionally, *de* and *bǎ* share similar prosodic and semantic constraints (Feng 2019; Li 2006; Liu 1997; Wang 2017). The prosodic constraint requires that the verb after *de*/*bǎ* must not be monosyllabic (2), and the semantic constraint rules out certain verb types from the *de*/*bǎ*-construction (3).[2]

(2) a. de^0 $van^{22}t\varepsilon\varepsilon^{44}$ fo^3pi^{35}/*fo^{323} $\mathfrak{Ʒi}^2\text{fiau}^{13}$, ni^{13} $\varepsilon i\varepsilon^0$ $t\varepsilon u^{35}$ nε^0 a^0 $\langle$W$\rangle$
       DE article   publish/*publish after   2SG want do   what Q
       'After publishing the article, what do you want to do?'
     b. bǎ wénzhāng fābiǎo/*fā   yǐhòu, nǐ xiǎng zuò shénme $\langle$M$\rangle$
       BA article   publish/*publish after 2SG want do what
       'After publishing the article, what do you want to do?'

(3) a. *$l\varepsilon^0$-$t\varepsilon\varepsilon^{44}$ de^0 $l\varepsilon^0$-fio^{31} $\varepsilon i^{53}\varepsilon y^{44}$/$t^h\varepsilon^{53}ji^{22}$-$\text{fio}^0$ $\langle$W$\rangle$
       old-Zhang DE old-Huang like/dislike-PFV
       Intended: 'Old-Zhang likes/dislikes old-Huang.'
     b. *lǎo-zhāng bǎ zhè gè gùshi xiāngxìn-le $\langle$M$\rangle$
       old-Zhang BA DEM CLF story believe-PFV
       Intended: 'Old-Zhang believed this story.'

In (2), the verb 'publish' has monosyllabic and disyllabic variants, i.e. *fopi* and *fo* in Wenzhounese versus *fābiǎo* and *fā* in Mandarin. Although all of them can be followed by *vantɕɛ*/*wénzhāng* 'article' in a canonical VO order, only the disyllabic ones are legitimate in (2). That is, verbs' syllabicity may affect the acceptability of the *de*/*bǎ*-construction. On the other hand, (3) shows that stative verbs such as *ɕiɕy* 'like', *tʰɛji* 'dislike', and *xiāngxìn* 'believe' are not allowed in the *de*/*bǎ*-construction at all (see Liu 1997 for details).

Section 2 provides a detailed comparison between Wenzhounese *de* and Mandarin *bǎ*. Since there is no formal analysis of Wenzhounese *de*, this paper fills this research

---

[2] I follow Leipzig glossing rules (www.eva.mpg.de/lingua/resources/glossing-rules.php) plus the following abbreviations: PRT - exclamative particle, RED - reduplication, SFP - sentence final particle, and SUB - subordinator.

gap within the framework of lexical functional grammar (LFG)[3]: Sect. 3 reviews previous LFG approaches to Mandarin *bǎ*, and Sect. 4 offers an LFG account for the *de*-construction. Section 5 concludes.

2 Comparing *de* and *bǎ*

This section elaborates on the similarities and disparities between *de* and *bǎ*, including the topicalization strategies (Sect. 2.1) and topic properties (Sect. 2.2) of post-*bǎ/de* NPs, the distribution of the universal quantifier (Sect. 2.3), the double-*de* construction (Sect. 2.4), the voice of post-*bǎ/de* verbs (Sect. 2.5), and the possibilities of admitting retained subjects/objects (Sect. 2.6).

2.1 Topicalization

As a topic-prominent or discourse-configurational language (Kiss 1995; Li and Thompson 1976; Xu 2015), Mandarin often marks the information-structural topic by having it occupy the sentence-initial position. This notwithstanding, it is generally impossible to topicalize the post-*bǎ* NP (McCawley 1992, p. 225). (4a) is a typical *bǎ*-sentence, where *bǎ* precedes the object NP *zhè-xiē píngguǒ* 'these apples'. As we can see in (4b), preposing the post-*bǎ* NP to the clause-initial topic position results in ungrammaticality, unless we use a resumptive pronoun after *bǎ* (4c).

(4)   a. wǒ bǎ zhè-xiē píngguǒ chī-le   ⟨M⟩
           1SG BA DEM- PL apple   eat-PFV
           'I ate these apples.'
      b. *[zhè-xiē píngguǒ]$_i$, wǒ bǎ Ø$_i$ chī-le   ⟨M⟩
           DEM- PL apple   1SG BA   eat-PFV
           Intended: 'These apples, I ate.'
      c. [zhè-xiē píngguǒ]$_i$, wǒ bǎ tā$_i$/tā-men$_i$ chī-le   ⟨M⟩
           DEM- PL apple   1SG BA 3SG/3- PL eat-PFV
           'These apples, I ate them.'

In contrast to Mandarin, the post-*de* NP in Wenzhounese can be topicalized, using either the gap strategy (5a) or resumption (5b).[4]

(5)   a. [ge^{212}-le^0 kai^0 zɿ^{31}ko^{53}]$_i$ nei^0, de^0 Ø$_i$ kʰo^{53} ta^0 ɕi^0 ⟨WSC⟩
           DEM- PL CLF issue   PRT, DE   put aside first
           'These issues, (let's) put them aside for the time being.'
      b. [ge^{212}-le^0 kai^0 zɿ^{31}ko^{53}]$_i$ nei^0, de^0 gi^0_i kʰo^{53} ta^0 ɕi^0 ⟨W⟩
           DEM- PL CLF issue   PRT, DE 3SG put   aside first
           'These issues, (let's) put them aside for the time being.'

---

[3] LFG is a non-derivational framework. In this paper, expressions with a derivational implication, e.g., extraction and topicalization, should be taken as pre-theoretical and do not reflect the theoretical assumptions underpinning LFG.

[4] For the sentences in (5), either a hortative (let's) or a command meaning is possible, and it is usually determined by pragmatics or contexts.

In (5b), the resumptive pronoun does not agree in number with its antecedent, while its Mandarin counterpart in (4c) shows optional number agreement. Note that number agreement is required in Wenzhounese when the object in an SVO sentence is topicalized (6), so the question is: why is it absent in the *de*-construction?

(6) ge^0 la^{13} kai^0 naŋ^{31}, ŋ^{13} za^{0}tsaŋ^{44} fɯ^0 ɕi^{53}ɕy^{44} gi^{31}-le^0 ⟨W⟩
     DEM two CLF person 1SG really   NEG like   3-PL
     'These two people, I really don't like them.'

The answer may be that there are two types of resumptive pronouns involved. Those with number agreement are ordinary pronouns, but those without are *processor resumptives*, which are not required by the grammar but "inserted where a gap would lead to ungrammaticality or processing difficulty" (Asudeh 2012, p. 41). For example, the underlined pronouns in (7) are processor resumptives, without which the sentences' acceptability will degrade.

(7) a. I'd like to meet the linguist that Mary couldn't remember if she had seen
        <u>him</u> before. (Sells 1984, p. 11)
     b. This is a donkey that I don't know where <u>it</u> lives. (Ferreira and Swets 2005,
        p. 271)

The insertion of processor resumptives is post-grammatical, i.e. they are "produced through incremental construction of locally well-formed structures" in the processing model (Asudeh 2012, p. 298). See also Engdahl (1982) for Swedish examples.

Similar to the pronouns in (7), the resumptive *gi* in the *de*-construction (glossed as GI hereinafter) helps alleviate island constraint violations. The data below show that extracting the post-*de* object from within a complex NP island impairs the sentences' acceptability.[5]

(8) a. ŋ^{13} ɕɔ^{35}te^0 [he^{35} kai^0 [de^0 pɛ^{0}ɕi^{53}-ɕɛ^{44} ta^{35}kʰe^0]$_{VP}$ gi^0 naŋ^{31}]$_{NP}$ ⟨W⟩
        1SG know DEM CLF DE secure-box open   SUB person
        'I know that person who opened the safe.'
     b. ?ŋ^{13} ɕɔ^{35}te^0 [he^{35} kai^0 pɛ^{0}ɕi^{53}-ɕɛ$_i^{44}$ [de^0 Ø$_i$ ta^{35}kʰe^0]$_{VP}$ gi^0 naŋ^{31}]$_{NP}$ ⟨W⟩
        1SG know DEM CLF secure-box DE   open   SUB person
        Intended: 'I know that person who opened the safe.'
     c. ?*pɛ^{0}ɕi^{53}-ɕɛ$_i^{44}$, ŋ^{13} ɕɔ^{35}te^0 [he^{35} kai^0 [de^0 Ø$_i$ ta^{35}kʰe^0]$_{VP}$ gi^0 naŋ^{31}]$_{NP}$ ⟨W⟩
        secure-box 1SG know DEM CLF DE   open   SUB person
        Intended: 'I know that person who opened the safe.'

The variance of acceptability suggests that information from non-syntactic module(s) plays a role in the interpretation of these sentences. Crucially, when the gap is replaced

---

[5] Native speakers' intuitions vary regarding these sentences. All my informants agreed that (8a) was fully acceptable. One of them found (8b) acceptable, one found it marginal, and the remaining two said it was unacceptable. As for (8c), half said it was marginally acceptable and the other half said it was unacceptable. My own native-speaker intuition is that (8a, b) are acceptable and (8c) is not. All but one informants agreed that the resumptive pronoun in (9) made the sentence more acceptable. To reflect different levels of acceptability, I use ?, ??, and ?* to mark sentences that are marginally acceptable, questionable, and marginally unacceptable, respectively.

by GI, the sentences become more acceptable, which is similar to (7) where the processor resumptives in English facilitate processing.

(9) a. ŋ^{13} çɔ^{35}te^0 [he^{35} kai^0 pɛ^{0}çi^{53}-çɛ$_i^{44}$ [de^0 gi$_i^0$ ta^{35}k^{h}e^0]$_{VP}$ gi^0 naŋ^{31}]$_{NP}$ ⟨W⟩
         1SG know DEM CLF secure-box DE GI open   SUB person
         'I know that person who opened the safe.'

     b. pɛ^{0}çi^{53}-çɛ$_i^{44}$, ŋ^{13} çɔ^{35}te^0 [he^{35} kai^0 [de^0 gi$_i^0$ ta^{35}k^{h}e^0]$_{VP}$ gi^0 naŋ^{31}]$_{NP}$ ⟨W⟩
         secure-box 1SG know DEM CLF DE GI open   SUB person
         'I know that person who opened the safe.'

In short, in Mandarin, topicalizing the post-*bǎ* NP must be accompanied by a resumptive pronoun. In Wenzhounese, either a gap or a resumptive pronoun GI is allowed. This GI seems to be a processor resumptive, an analysis for which I present further support in Sect. 2.3.

2.2 Topic properties

We have seen in the previous section that the post-*de* NP can be topicalized. This section shows that even setting aside topicalization, the in-situ NP manifests topic properties.

Listing the following properties of topic in Mandarin Chinese, Tsao (1987) demonstrates that the post-*bǎ* NP exhibits almost all these properties, except for (10f), which is difficult to test in the *bǎ*-construction (p. 11).[6]

(10) Topic properties according to Tsao (1987, p. 4)

     a. Topic invariably occupies the S-initial position of the first sentence in a topic chain.
     b. Topic can optionally be separated from the rest of the sentence in which it overtly occurs by one of the four pause particles: *a* (*ya*), *na*, *me*, and *ba*.[7]
     c. Topic is always definite or generic.
     d. Topic is a discourse notion: it may, and often does, extend its semantic domain to more than one sentence.
     e. Topic is in control of the pronominalization or deletion of all the coreferential NPs in a topic chain.
     f. Topic, except in sentences where it is also subject, plays no role in such processes as reflexivization, passivization, Equi-NP deletion and imperativization.

---

[6] An anonymous reviewer points out that the post-*bǎ* NP cannot satisfy (10a) as it is not S-initial. According to Tsao (1987, p. 11), the post-*bǎ* NP "occupies the S-initial position of the first S in a *bǎ* topic chain." This means that the NP need not be absolutely sentence-initial, as long as it is initial in a topic chain. A Wenzhounese example is provided in (11), in which the post-*de* NP *vu* 'house' is initial in the clauses headed by *ts^{h}a-tç^{h}i* 'painted' and *ma-tç^{h}y-fiɔ* 'sold'.

[7] These "pause particles" are also known as topic markers (Xu 2015, p. 395).

The post-*de* NP in Wenzhounese shares most of these topic properties.[8] For example, it can head a topic chain (11) and it can be followed by a topic marker, i.e. *ne* in (12).

(11) lɛ^0-tɕɛ^{44} de^0 vu^{212} ta^{53}sɛ^{53}-fiɔ^0, jau^{13}-tsʰa^{323} tsʰa^{323}-tɕʰi^0, ga^{323}
old-Zhang DE house clean-PFV   oil-paint$_N$ paint$_V$-up   then
ma^{22}-tɕʰy^{30}-fiɔ^0 ⟨W⟩
sell-out-PFV
'Old Zhang had the house cleaned, painted, and then sold.'

(12) ŋ^{13}-le^0 de^0 jau^{35}-le^0 zəŋ^{13}tɕi^{323} ne^0, kʰɔ^0 da^{13}ko^{33}naŋ^{21} ko^{35}-ʔe^0-ko^{35}
1- PL   DE some-PL plot   PRT to   everyone   tell-one-tell
ba^{35} ge^0 ⟨WSC⟩
SFP SFP
'We have already talked a little about some of the plots with everybody.'

Moreover, the post-*de* NP can establish a whole-part or possessor-possessee relation with a following NP, just as sentence-initial topics may do (Tsao 1987, pp. 18–20).

(13) a. ŋ^{13} de^0 [ke^{44}]$_{whole}$ tɕʰi^0 [la^{13} sɔ^{53}]$_{part}$ d-ɔ^{31} ⟨W⟩
1SG DE orange   eat two segment inside-PFV
'I ate two segments of the orange.'
b. ɲi^{13} de^0 [ge^0 bu^0 tsʰo^{44}]$_{possessor}$ [ko^{35}ka^{323}]$_{possessee}$ ma^0 bi^{212}-le^0
2SG DE DEM CLF car   price   sell cheap-bit
ma^0 ⟨W⟩
SFP
'Please sell this car at a lower price.'

Summarizing, the post-*bǎ/de* NPs exhibit topic properties outlined by Tsao (1987), suggesting that they should be analyzed as a topic at some level of representation.

2.3 Distribution of quantifiers

In Wenzhounese, the distribution of the quantifier is sensitive to whether the post-*de* NP is a resumptive pronoun or a regular (pro)nominal. To begin with, consider the universal quantifier *dōu* 'all' in Mandarin. In the *bǎ*-construction, *dōu* must follow post-*bǎ* NP to take scope over it (cf. Huang et al. 2009, p. 180ff). In (14), therefore, *dōu* cannot precede *bǎ*. The quantifier at issue is underlined.

(14) wǒ (*dōu) bǎ zhè-xiē píngguǒ (dōu) chī-le   ⟨M⟩
1SG (*all) BA DEM- PL apple   (all) eat-PFV
'I ate all these apples.'

---

[8] Topics and post-*de* NPs need not be definite or generic in Wenzhounese (cf. 10c), as expressions like *za-ke paŋ səɯ* (十几本书) 'about ten books' can occupy these positions.

When the post-*bǎ* position is filled by a resumptive pronoun, the quantifier still occupies the same position, as (15) shows. Note that the quantifier cannot immediately follow the topic.

(15) [zhè-xiē píngguǒ]$_i$, wǒ (*dōu) bǎ tā$_i$/tā-men$_i$ (dōu) chī-le   ⟨M⟩
  DEM- PL apple 1SG (*all) BA 3SG/3- PL (all) eat-PFV
  'These apples, I ate them all.'

The universal quantifier in Wenzhounese patterns with its Mandarin counterpart when there is no topicalization. That is, when the post-*de* NP stays in-situ, the quantifier must follow this NP, and thus also follows *de*.

(16) ŋ^{13} (*ʔo^{323}) de^0 lɔ^{25}dʑa^{212} /gi^{31}-le^0 (ʔo^{323}) tɛ^{53}-kʰɔ^0 ⟨W⟩
  1SG (*all) DE trash /3- PL (all) dump-away
  'I dumped all of the trash/them.'

However, when the post-*de* NP is topicalized and leaves a gap, the quantifier preferably precedes *de* (17a).[9] Crucially, when GI occupies the post-*de* position (17b), the quantifier also needs to precede *de*, in opposition to the non-resumptive pronoun in (16). This contrasts with the *bǎ*-construction in Mandarin, where the quantifier uniformly follows the post-*bǎ* NP.

(17) a. lɔ^{25}dʑa$_i^{212}$, ŋ^{13} (ʔo^{323}) de^0 Ø$_i$ ($^{??}$ʔo^{323}) tɛ^{53}-kʰɔ^0 ⟨W⟩
    trash, 1SG (all) DE ($^{??}$all) dump-away
    'The trash, I dumped it all.'
    b. lɔ^{25}dʑa$_i^{212}$, ŋ^{13} (ʔo^{323}) de^0 gi^0 ($^{??}$ʔo^{323}) tɛ^{53}-kʰɔ^0 ⟨W⟩
    trash, 1SG (all) DE GI ($^{??}$all) dump-away
    'The trash, I dumped it all.'

This gap-like property of GI is confined to the *de*-construction, if we compare it with topicalization in SVO sentences. When using the gap strategy, we can insert a quantifier between the subject and the predicate (18a). The placement of the quantifier is similar to (17), i.e. immediately after the subject. The quantifier cannot be used when there is no topicalization (18b) or when there is an "ordinary" resumptive pronoun with number agreement. This differs from *de*-sentences like (17b) that contains GI, as they allow the quantifier to intervene between the subject and *de*.

(18) a. ge^0 la^{13} kai^0 naŋ^{31}, ŋ^{13} ʔɔ^{323} ɕi^{53}ɕy^{44} Ø ⟨W⟩
    DEM two CLF person 1SG all like
    'These two people, I like both of them.'
    b. *ŋ^{13} ʔɔ^{323} ɕi^{53}ɕy^{44} ge^0 la^{13} kai^0 naŋ^{31} ⟨W⟩
    1SG all like DEM two CLF person
    Intended: 'I like both of these two people.'
    c. *ge^0 la^{13} kai^0 naŋ^{31}, ŋ^{13} ʔɔ^{323} ɕi^{53}ɕy^{44} gi^{31}-le^0 ⟨W⟩
    DEM two CLF person 1SG all like 3-PL
    Intended: 'These two people, I like both of them.'

---

[9] My informants consistently preferred the quantifier to precede *de*, although they still found it grammatical for the quantifier to follow the gap or GI.

In short, the resumptive pronoun in (18) shows number agreement and blocks quantifier float, suggesting that it plays an active role in the grammar. GI in (17) does not show number agreement and triggers quantifier float, just like a gap. This can be explained if the post-*de* GI is a processor resumptive; that is, GI is not licensed by competence but by performance, so it does not interact with the grammar (Asudeh 2011).

Finally, there are also phonological differences between GI and "true" resumptives. The latter, as in (18), may surface with their underlying tone, e.g., [gi^{31}] '3SG' and [gi^{31}-le^0] '3PL'. The processor resumptive GI, as in (17), obligatorily undergoes tonal neutralization (indicated by "0") and optionally undergoes lenition for some speakers: [gi^0] → [ji^0]. In other words, the processor resumptive is prosodically weaker than the true resumptive.

2.4 The double-*de* construction

Another disparity between *bǎ* and *de* is that a *de*-sentence can have two *de*s (which I will call the double-*de* construction), whereas a *bǎ*-sentence can only have one *bǎ*.

(19) a. tsəŋ^{53}fuɪ^{35} **de**^{0} [ge^{212}-le^0 tɕy^{0}ŋo^{31} jɛ^{31}]$_i$ ʔo^{212} **de**^{0} (gi^0_i)
         government DE DEM- PL brick   wall all   DE (GI)
         tsʰa^0-tsʰa^{323}-fɪɔ^0 ⟨WSC⟩
         demolish-RED- PFV
         'The government demolished all these brick walls.'

   b. zhèngfǔ   **bǎ** zhè-xiē zhuānwǎ qiáng dōu (*bǎ tā)
         government BA DEM- PL brick   wall all (*BA 3SG)
         chāi-le   ⟨M⟩
         demolished-PFV
         'The government demolished all these brick walls.'

(20)   ... gi^{31} tɕu^0 nɔ^{22}ku^{35}-tɕʰi^0 nei^0, zou^{212} **de**^{0} [tsʰo^{44} gi^0 pu^{33}lei^{21}]$_i$
         3SG be.made upset-up   PRT, then   DE vehicle POSS window
         de^0 Ø$_i$ hɛ^0-hɛ^{53}-fɪɔ^0 ⟨WSC⟩
         DE   smash-RED- PFV
         'After s/he was made upset (by taking the bus), s/he smashed the bus's window.'

In (19a) and (20), there are two occurrences of *de*, the second of which can be followed by either a gap or GI. By contrast, a parallel double *bǎ*-construction such as (19b) is ungrammatical in Mandarin.

An important restriction on the double-*de* construction is that the second *de* can only be followed by a gap or GI, but not a full NP (21a). The first post-*de* NP usually does not topicalize so as to avoid two consecutive tokens of *de* (21b), but topicalization is possible when there is an intervening verb (21c). These generalizations are summarized in (22).

(21) a. *gi^{31} **de**^{0} pɛ^{0}ɕi^{53}-ɕɛ^{44} **de**^{0} pɛ^{0}ɕi^{53}-ɕɛ^{44} ta^{35}kʰe^{44} ⟨W⟩
         3SG DE secure-box DE secure-box open
         Intended: 'S/he opened the safe.'

b. *$p\varepsilon^0\varsigma i^{53}$-$\varsigma\varepsilon_i^{44}$, gi^{31} **de^0** $\emptyset_i$ **de^0** (gi_i^0) $ta^{35}k^he^{44}$ ⟨W⟩
  secure-box 3SG DE   DE (GI) open
  Intended: 'The safe, s/he opened it.'

c. $p\varepsilon^0\varsigma i^{53}$-$\varsigma\varepsilon_i^{44}$, gi^{31} **de^0** $\emptyset_i$ $\varsigma\varepsilon^0$ $b\mathit{o}^{13}f\mathit{o}^{323}$ **de^0** (gi_i^0) $ta^{35}k^he^{44}$ ⟨W⟩
  secure-box 3SG DE   think method DE (GI) open
  'The safe, s/he thought of a way to open it.'

(22) a. Subject **de** Object **de** (GI) V
   b. *Subject **de** Object **de** Object V
   c. *Topic$_i$ Subject **de** $\emptyset_i$ **de** (GI) V
   d. Topic$_i$ Subject **de** $\emptyset_i$ V$_1$ **de** (GI) V$_2$

Thus, there is an asymmetry in the double-*de* construction: while the first *de* can be followed by a full NP, the second *de* cannot.

2.5 Unmarked passives

In Mandarin, passivization can be marked by the morpheme *bèi*, as in (23a). Tan (1991, Ch.3) uses tests like reflexive binding, adjunct control, and imperatives to show that (23b) is also a passive sentence, albeit morphologically unmarked.

(23) a. lisì bèi (lǐngdǎo) xiàfàng le   ⟨M⟩
     Lisi PASS (leader) demote SFP
     'Lisi was demoted (by the leader).'
   b. lǐsì xiàfàng le   ⟨M⟩
     Lisi demote SFP
     'Lisi was demoted.'

Bender (2000, Sect. 7.2) argues that in the core cases of the *bǎ*-construction (i.e. those without "retained objects", to be discussed in Sect. 2.6), the embedded verb is also an unmarked passive. The questions are (i) whether unmarked passives are attested in Wenzhounese and (ii) whether the post-*de* verb can be an unmarked passive.

It turns out that unmarked passives are widely used in Wenzhounese complex predicates. As Pan (1997) observes, resultative-verb compounds (RVCs) in Wenzhounese must be preceded by their objects, despite the canonical SVO word order.

(24) a. *$l\varepsilon^0$-$t\varsigma\varepsilon^{44}$ ςy^0-$vu^{44}t\varsigma y^{44}$-$f\mathit{o}^0$ p^hai^{53} ⟨W; SVO⟩
     old-Zhang cook-burn-PFV   dish
     Intended: 'Old Zhang overcooked the dish.'
   b. $l\varepsilon^0$-$t\varsigma\varepsilon^{44}$ p^hai^{53} ςy^0-$vu^{44}t\varsigma y^{44}$-$f\mathit{o}^0$ ⟨W; SOV⟩
     old-Zhang dish cook-burn-PFV
     'Old Zhang overcooked the dish.'
   c. $l\varepsilon^0$-$t\varsigma\varepsilon^{44}$ de^0 p^hai^{53} ςy^0-$vu^{44}t\varsigma y^{44}$-$f\mathit{o}^0$ ⟨W; S*de*OV⟩
     old-Zhang DE dish cook-burn-PFV
     'Old Zhang overcooked the dish.'

In what follows, I provide evidence that the ostensible SOV order is actually TSV ("T" stands for Topic), in which the RVC is an unmarked passive. First, the sentence-initial NP in (25b) cannot bind the subject-oriented reflexive zl, while the one in the *de*-construction can.[10] This suggests that *lɛ-tɕɛ* 'old Zhang' bears the subject function in the *de*-sentence (25b) but does not in the ostensible SOV sentence (25a).

(25)　a.?*$lɛ^0$-$tɕɛ^{44}_i$　zl^{22}_i-da^{22}　p^hai^{53}　$ɕy^0$-$vu^{44}tɕy^{44}$-$fiɔ^0$ ⟨W⟩
　　　　old-Zhang self-place dish　cook-burn-PFV
　　　　Intended: 'Old Zhang overcooked his own dish.'
　　b.　$lɛ^0$-$tɕɛ^{44}_i$　de^0 zl^{22}_i-da^{22}　p^hai^{53}　$ɕy^0$-$vu^{44}tɕy^{44}$-$fiɔ^0$ ⟨W⟩
　　　　old-Zhang DE self-place dish　cook-burn-PFV
　　　　'Old Zhang overcooked his own dish.'

Second, subject-oriented adverbs like *dede-naŋ* 'deliberately' are incompatible with ostensible SOV sentences (26a) but compatible with *de*-sentences (26b).

(26)　a.　*$lɛ^0$-$tɕɛ^{44}$　de^0de^{22}-$naŋ^0$　p^hai^{53}　$ɕy^0$-$vu^{44}tɕy^{44}$-$fiɔ^0$ ⟨W⟩
　　　　old-Zhang deliberate-look dish　cook-burn-PFV
　　　　Intended: 'Old Zhang deliberately overcooked the dish.'
　　b.　$lɛ^0$-$tɕɛ^{44}$　de^0de^{22}-$naŋ^0$　de^0 p^hai^{53}　$ɕy^0$-$vu^{44}tɕy^{44}$-$fiɔ^0$ ⟨W⟩
　　　　old-Zhang deliberate-look DE dish　cook-burn-PFV
　　　　'Old Zhang deliberately overcooked the dish.'

These disparities can be accounted for if the RVCs are unmarked passives that only subcategorize for a PATIENT subject. The sentence-initial NPs in these sentences, e.g., *lɛ-tɕɛ* 'old Zhang' in (25a) and (26a), are topics whose semantic role is AGENT. In short, the ostensible SOV order is actually TSV, with a verb in the unmarked passive.

The same reasoning can be applied to verbs suffixed by the perfective marker *-fiɔ* or the multifunctional aspect marker *-tɕʰi* (e.g., *tso-tɕʰi* 'install-up' and *tɕɔ-tɕʰi* 'wear-up').[11] While the corresponding simplex verbs (i.e. monomorphemic verbs without the aspectual marker) are atelic and require the SVO order, the V-ASP verbs pattern with RVCs in terms of telicity, word order, reflexive binding, and adverbial modification. In other words, these complex verbs are also unmarked passives.

Tan (1991, Sect. 3.5.1) has shown that in Mandarin, only telic verbs can passivize. In Wenzhounese, RVCs and V-ASP verbs are also telic, and thus they form a natural class for (unmarked) passivization. Crucially, all these unmarked passive verbs can enter the *de*-construction. Given that passivization is a lexical rule (Bresnan 1982b) and the *de*-construction is a syntactic construction, the syntax cannot alter the voice of a verb as per the principle of lexical integrity (Bresnan and Mchombo 1995). Consequently, a passive verb must remain passivized in the *de*-construction.

To sum up, this section has argued that RVCs and V-ASP forms in Wenzhounese are unmarked passives, which explains why these complex verbs cannot take a post-verbal object in an otherwise SVO language. By contrast, their Mandarin counterparts

---

[10] One informant found (25a) marginally acceptable, so I did not mark it with an asterisk.

[11] Pan (2002) offers a detailed discussion of the aspect and mood in Wenzhounese.

can still bear the active voice and take post-verbal objects, although it is usually their passivized version that enters the *bǎ*-construction (Bender 2000).

2.6 Retained objects and retained subjects

Post-*bǎ* verbs can be unmarked passives, but are not necessarily so. Some of them can take an additional object (Hsueh 1989). In (27), for instance, *bǎ* is followed by the NP *júzi* 'orange', and the verb *bō* 'peel' is followed by another NP *pí* 'skin'. Such data are problematic for analyses that treat *bǎ* as an object marker (e.g., Peyraube and Wiebusch (2021)) because an *ad hoc* rule is required to increase the valence of *bō* 'peel'.

(27) tā   bǎ júzi   bō-le   **pí**   ⟨M⟩
3SG BA orange peel-PFV skin
'S/he peeled the orange.'

The object of the post-*bǎ* verb, in addition to the one that *bǎ* allegedly object-marks, is known as the retained object (Bender 2000, Sect. 3.2). As Tsao (1987, p. 19) observes, the post-*bǎ* NP and the retained object are either in a possessor-possessee relation or a whole-part relation. We have seen in (13) that such retained objects are also allowed in the *de*-construction.

Moreover, the post-*de* verb may take a subject, alongside the post-*de* NP. What I shall refer to as "retained subject", to the best of my knowledge, is not reported in studies of the *bǎ*-construction. In (28), the first line indicates the underlying tone values from *seiku bei* 'watermelon peel'. The underlying tones may change in connected speech, a process known as tone sandhi (see, e.g., Bao 2011; Chen 2000; Zhengzhang 2008). For (28), the surface tones can be realized in two ways. The first possibility is for *seiku bei* to undergo trisyllabic tone sandhi (Chen 2000, p. 480), which is a lexical phonological rule operative on compounds. This is shown in (28a) and the brackets delimit the tone sandhi domain. As such, *bei* is part of the compound *seiku bei*, which as a whole serves as the post-*de* NP. Another possibility (28b) is for *seiku* to undergo disyllabic tone sandhi (also a lexical rule), to the exclusion of *bei*. In this case, there is a noticeable pause after *seiku*, indicating a prosodic boundary and suggesting that *seiku* and *bei* belong to different syntactic phrases if there is a general alignment between syntactic and prosodic constituents (Selkirk 1986, 2011).

(28) Underlying tones: de^0 sei^{44}ku^{44} bei^{31} tɕʰi^{323}-d-ɔ^0 ⟨W⟩
      Gloss: DE   watermelon   peel   eat-inside-PFV
      a: (sei^{35} ku^{22} bei^{22})
      b: (sei^{44}ku^{44})   (bei^{31})
Translation: 'eat up the watermelon peel'

Further support for the syntactic phrase boundary comes from questionability and topicalization. For the question in (29a), only the reading in (28a) is legitimate; but if the question is (29b), only (28b) is a possible answer.

(29)   a.   gi^{31} de^0 [n̠ɛ^{31} mø^{53}z̩]^{212}]$_{NP}$ tɕʰi^{323}-d-ɔ^{22} a^0 ⟨W⟩
             3SG DE what thing   eat-inside-PFV Q
             'What did s/he eat up?'
        b.   gi^{31} de^0 [n̠ɛ^{31} mø^{53}z̩]^{212}]$_{NP}$ [bei^{31}]$_{NP}$ tɕʰi^{323}-d-ɔ^{22} a^0 ⟨W⟩
             3SG DE what thing   skin   eat-inside-PFV Q
             'What did s/he eat up the skin of?'

In terms of topicalization, we can see in (30) that *seiku* can be topicalized, which is
impossible if *seiku* is a possessor of *bei* within the same NP because the presence of
the possessive clitic *gi* incurs ungrammaticality.

(30)   sei^{44}ku$_i^{44}$ (*=gi^0),   ŋ̩^{13} de^0 Ø$_i$ bei^{31} tɕʰi^{323}-d-ɔ^{22} ⟨W⟩
        watermelon (*=POSS) 1SG DE   peel eat-inside-PFV
        'As for the watermelon, I ate up its peel.'

Another instance of the retained subject is *kʰoŋ duɔ* 'air conditioner' in (31). It
is a subject because the verb *tso-tɕʰi* 'install-up' is passivized (Sect. 2.5) and cannot
take an object. Nor does this verb subcategorize for the argument *kosa tei* 'inside
the classroom', which therefore must be introduced by *de*. This argument contains a
nominal *kosa* 'classroom' and a localizer *tei* 'inside', but it is still nominal in nature
(Li 2019; Nie and Liu 2021).[12]

(31)   gi^{31}-le^0 de^0 [ko^{35}sa^{323} tei^0]$_{NP}$ kʰoŋ^{22}duɔ^{35} ʔo^{323} tso^{44}-tɕʰi^0 ba^{13} ⟨W⟩
        3- PL   DE classroom inside air.conditioner all   install-up SFP
        'They've installed the classrooms with air-conditioners.' (Xu 2019, p. 118)

Note that Mandarin also admits a [nominal + localizer] NP after *bǎ*, as in (32).[13]
The difference is that *tián-mǎn* 'fill-full' is an active verb whose post-verbal NP is its
object.

(32)   bǎ [fángjiān lǐ]$_{NP}$ tián-mǎn gèzhǒng   shíyòng de   dōngxi ⟨M⟩
        BA room   inside fill-full all.kinds.of useful   SUB thing
        'stuff the room with all kinds of useful things'

The main point is that both the *bǎ-* and the *de*-constructions can include a retained
object, but only the latter can include a retained subject. This is a consequence of
unmarked passivization of telic verbs in Wenzhounese (Sect. 2.5), due to which the
object of a transitive verb becomes the subject of the corresponding passivized verb.
The existence of retained subjects indicates a structural difference between the post-
bǎ and post-*de* verb: the former heads a VP but the latter an IP (Sect. 4.1). Also, the
grammatical function associated with this IP must not be an open function with a

---

[12] I thank an anonymous reviewer for suggesting the references.

[13] This sentence comes from the CCL Corpus of Chinese Texts (Zhan et al. 2003), but its acceptability
varies among speakers. I informally asked five native Mandarin speakers to rate this sentence against a five-
point Likert Scale. Two of them found it acceptable (4 and 5, respectively), one rated 3, and the remaining
two rated 2, considering it virtually unacceptable. It is unclear whether the variation in acceptability is due
to grammatical or extra-grammatical factors. By contrast, the four Wenzhounese informants considered a
[nominal + localizer] NP in the *de*-construction fully acceptable.

missing subject (XCOMP in LFG) because of the existence of retained subjects (Sect. 3.2).

2.7 Summary

Section 2 has shown that both Wenzhounese *de* and Mandarin *bǎ* allow retained objects and often take passivized verbs that are morphologically unmarked. Also, the post-*de/bǎ* NPs in both languages have properties typical of topics. Despite these similarities, they differ in the following respects:

(33)   a. In Mandarin, resumption is the only way to topicalize the post-*bǎ* NP. By contrast, Wenzhounese can use either resumption or the gap strategy.
   b. The resumptive pronoun in Mandarin shows (optional) number agreement, but its Wenzhounese counterpart GI does not.
   c. In the *bǎ*-construction, the quantifier uniformly follows the post-*bǎ* NP. In Wenzhounese, the quantifier follows the post-*de* NP when there is no topicalization; the quantifier precedes *de* when there is topicalization.
   d. There is a double *de*-construction which is unattested in Mandarin.
   e. Only Wenzhounese can contain a retained subject in the *de*-construction.

I have argued above that GI is a processor resumptive that is licensed by the post-grammatical processing model, hence its lack of number agreement (33b). Note also that (33e) can be readily explained by the extensive use of unmarked passives in Wenzhounese. Still, a full analysis of *de* needs to capture the remaining differences. I offer such an analysis in Sect. 4. Before that, a review is necessary of the previous investigation of Mandarin *bǎ*, which may be relevant to the analysis of *de*.

3 Previous analyses of *bǎ*

There is only non-formal, descriptive work on Wenzhounese *de*, e.g., Chen (2010), Lin (2019, Sect. 3.2), and Zhengzhang (2008, pp. 245–246). I will therefore focus on Mandarin *bǎ*, whose synchronic, diachronic, and formal properties have been extensively studied (see, e.g., Huang et al. (2009) and Sun (2015) for an overview). This notwithstanding, not even *bǎ*'s part-of-speech is settled. There are claims that it is a preposition (Her 1990; McCawley 1992), a verb (Hashimoto 1971; Bender 2000), or a functional category (Sybesma 1999; Huang et al. 2009; Li 2006). Most of this work (including recent papers such as Shu 2018; Sun 2018; Wang 2017; Zhao 2021) is couched in a Chomskyan approach to syntax, and may be incompatible with the non-derivational, strictly modular framework of LFG. For example, that *bǎ* projects a functional category, e.g., baP or VoiceP, is generally not accepted in LFG, because functional categories "are only warranted when a particular functional feature is associated with a structural position" (Bögel et al. 2018, p. 110).

My main focus in the following is LFG approaches to the *bǎ*-construction, specifically Her (1990), Bender (2000), and Her (2009).[14] Nevertheless, the points that I make will be as theory-neutral as possible, so they should easily extend to other frameworks. In Sect. 3.1, I introduce the core concepts of LFG. Section 3.2 discusses previous analyses of *bǎ* and their compatibility with Wenzhounese *de*.

3.1 Introducing lexical functional grammar

As the name suggests, LFG is a *lexical* theory, where "regularities across classes of lexical items are part of the organization of a richly structured lexicon" (Dalrymple et al. 2019, p. 3). The theory is *functional* because grammatical functions like subject and object are theory primitives. For details of LFG formalisms and their implementation, the reader may find useful Falk (2001), Bresnan et al. (2016), Börjars et al. (2019), and Dalrymple et al. (2019). Shorter introductions include Carnie (2013, Ch.16), Müller (2020, Ch.7), and Sells (2013).

LFG is a modular theory that posits various domain-specific linguistic components.[15] One such component is f(unctional)-structure. F-structure is a syntactic structure that encodes grammatical functions (GFs) such as subject and object, along with grammatical features like number, tense, and gender. Relevant for this paper are the following GFs: SUBJ(ect), OBJ(ect), COMP(lement), XCOMP(lement), and ADJ(unct). Among them, XCOMP and COMP are clausal functions. The former is an open function lacking an internal subject, while the latter is a closed function containing an internal subject (Dalrymple et al. 2019, p. 29). To illustrate, (34b) is the f-structure for (34a).

(34) a. ŋ   tɕiaŋjɛ tɕʰi seiku   ⟨W⟩
          1SG often   eat watermelon
          'I often eat watermelons.'

  b. $\begin{bmatrix} \text{PRED} & \text{'eat⟨SUBJ,OBJ⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{NUM} & \text{SG} \\ \text{PERS} & 1 \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'watermelon'} \end{bmatrix} \\ \text{ADJ} & \left\{ \begin{bmatrix} \text{PRED} & \text{'often'} \end{bmatrix} \right\} \end{bmatrix}$

We can see in (34b) that f-structure is an unordered attribute-value matrix, so it disregards linear order and constituency. These pieces of information are encoded in a distinct but related syntactic module called c(onstituent)-structure. It is organized as per X-bar theory, but a principle of Economy (Bresnan et al. 2016, p. 90) prunes out empty nodes and non-branching intermediate nodes. For example, (35) is the c-structure for (34a).

---

[14] Li (2019) is one of the most recent LFG accounts for *bǎ*, but it deals with mapping theta roles to grammatical functions and hence is irrelevant to this paper. Also, I will not elaborate on Kit's (1998) LFG analysis of *bǎ* because it is incompatible with some fundamental assumptions of LFG (cf. Bresnan 1982a, p. 372). For an up-to-date overview of LFG approaches to Sinitic languages, see Lam et al. (to appear).

[15] See, e.g., Coltheart (1999), Fodor (1983), and Scheer (2020) for modularity and domain specificity.

(35)
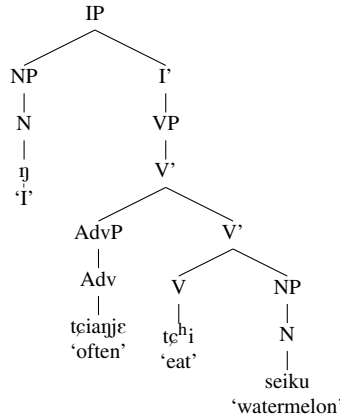
C-structures are licensed by annotated phrase structure rules, which specify how c-structural nodes are mapped to f-structures. For example, in Mandarin and Wenzhounese, but not necessarily other languages, the specifier of IP is associated with SUBJ, while the complement of V is associated with OBJ. We can express such associations by (36):

(36)   a.   IP   →   NP   I'
                        ($\uparrow$ SUBJ) = $\downarrow$ $\uparrow$ = $\downarrow$
    b.   V'   →   V   NP
                      $\uparrow$ = $\downarrow$ ($\uparrow$ OBJ) = $\downarrow$

The arrows in (36) are *metavariables* (Falk 2001, p. 69): $\uparrow$ refers to the f-structure of the mother node and $\downarrow$ to the f-structure of the current node. In (36a), $\uparrow$ = $\downarrow$ under I' means that the mother node of I', i.e. IP, maps to the same f-structure as I' does. The annotation under NP says that NP's f-structure is the SUBJ of IP's f-structure. Similarly, ($\uparrow$ OBJ) = $\downarrow$ in (36b) says "my mother's object is me."

These annotated phrase structures relate c-structure to f-structure, but they give only partial information about f-structure and are designed to work in consort with lexical entries.

(37)   a.   *$t\wideparen{\varsigma}^h i$*   V   ($\uparrow$ PRED) = 'eat⟨SUBJ,OBJ⟩'
    b.   *seiku*   N   ($\uparrow$ PRED) = 'watermelon'

(37) contains partial lexical entries for *$t\wideparen{\varsigma}^h i$* and *seiku*, which list their syntactic categories and f(unctional)-descriptions.[16] For example, (37a) specifies that *$t\wideparen{\varsigma}^h i$* is a two-place verb, contributing to f-structure the value of a PRED(icate) with a valency of two (two arguments appear in the bracketed argument list).

In short, LFG posits two structures to represent different aspects of syntactic information. F-structure records grammatical functions and features, while c-structure delineates linear order and constituency. Although these structures are independent, they are mutually constrained by lexical information and annotated phrase structure

---

[16] I only include morphosyntactic information here, but the lexical entry also contains phonological and semantic information. See Bögel (2015) and Dalrymple et al. (2019).

rules. We are now equipped with enough formal details to evaluate LFG analyses of the *bǎ*-construction.

3.2 The *bǎ*-construction in LFG

In this section, I compare the proposals of Her (1990, 2009), and Bender (2000). Although these are LFG analyses, most of my points will be general and not specific to LFG. I will show that none of these accounts can readily capture the *de*-construction in Wenzhounese, given the differences between *de* and *bǎ* summarized in (33).

To begin with, Her (1990) proposes that *bǎ* is a preposition, taking the post-*bǎ* NP as its complement. This PP is subcategorized for by the "main verb" (i.e. the post-*bǎ* verb). There are three problems with his analysis. First, most verbs that can enter the *bǎ*-construction have a non-*bǎ* counterpart, as exemplified in (38). If the alleged PP is an argument of the main verb, then these verbs need distinct lexical entries for *bǎ*- and non-*bǎ*-sentences. This is not economical, given that such verbs are abundant in Mandarin.

(38)   a.   wǒ bǎ shuǐ hē-le   ⟨M⟩
             1SG BA water drink-PFV
             'I drank the water.'
       b.   wǒ hē-le   shuǐ ⟨M⟩
             1SG drink-PFV water
             'I drank water.'

Second, Her's analysis predicts that the matrix subject must be selected by the main verb. However, there are cases, both in Mandarin and Wenzhounese, where the main verb does not subcategorize for the matrix subject. In both examples below, the RVC meaning 'fall asleep as a result of watching' is an intransitive verb subcategorizing for an EXPERIENCER subject, but the matrix subject 'this movie' is incompatible with this theta role. Therefore, the matrix subject's subjecthood and theta role should come from *bǎ/de* (Bender 2000, p. 110).

(39)   a.   zhè diànyǐng bǎ tā-men dōu kàn   shuìzháo le   ⟨M⟩
             DEM movie   BA 3- PL   all watch fall.asleep PFV
             'Watching this movie made all of them fall asleep.'
       b.   ge⁰   di⁰jaŋ³⁵ de⁰ gi³¹-le⁰ ʔo³²³ tsʰ\h ⁰   kʰy⁵³dzɔ⁰ ⟨W⟩
             DEM movie   DE   3- PL   all   watch   fall.asleep
             'Watching this movie made all of them fall asleep.'

Finally, that *bǎ* is a preposition is not well justified. Ostensible supporting evidence comes from McCawley (1992), who lists five tests to distinguish prepositions from verbs and concludes that *bǎ* is a preposition. However, Bender (2000, p. 137) demonstrates that these tests actually "support an analysis of *bǎ* as a verb". Of particular interest is McCawley's (1992, p. 220) observation that "objects of Vs usually can undergo extraction or deletion, while objects of Ps are less free in allowing extraction or deletion."

(40) a. zhāngsān$_i$, [wǒ méi pīpíng Ø$_i$] ⟨M⟩
          Zhangsan 1SG NEG criticise
          'Zhangsan, I didn't criticise.'

      b. wǒ jiā [lí nà gè gōngyuán]$_{PP}$ hěn yuǎn ⟨M⟩
          1SG house from DEM CLF park very far
          'My house is very far from that park.'

      c. *[nà gè gōngyuán]$_i$, wǒ jiā [lí Ø$_i$]$_{PP}$ hěn yuǎn ⟨M⟩
          DEM CLF park 1SG house from very far
          Intended: 'That park, my house is very far from.'

      d. *[nà gè júzi]$_i$, tā [bǎ Ø$_i$] bō-le pí ⟨M⟩
          DEM CLF orange 3SG BA peel-PFV skin
          Intended: 'That orange, he peeled.' (McCawley 1992, pp. 221, 225)

We can see above that the object of Vs can be topicalized (40a), while the object of the preposition *lí* 'from' cannot (40b–c). The post-*bǎ* NP cannot be topicalized either (40d), suggesting that *bǎ* is a preposition. However, extracting the post-*de* NP is allowed in Wenzhounese (see (5)), but extraction from within a PP is nevertheless prohibited (41):

(41) a. ŋ^{13} [mu^0 y^{44}tsou^{44}]$_{PP}$ tsau^{35} ⟨W⟩
          1SG to Wenzhou walk
          'I'm going to Wenzhou.'

      b. *y^{44}tsou$^{44}_i$, ŋ^{13} [mu^{212} Ø$_i$]$_{PP}$ tsau^{35} ⟨W⟩
          Wenzhou 1SG to walk
          Intended: 'Wenzhou, I'm going to.'

The contrast between (5) and (41b) suggests that Wenzhounese *de* is not a preposition. It also provides indirect evidence that the ban of extraction from the post-*bǎ* position may result from other factors than *bǎ* being a preposition. One such factor can be found in Bender (2000), who offers an independent reason for the ban on post-*bǎ* NP's extraction.

In Bender's (2000) view, *bǎ* is a verb. Observing that the post-*bǎ* NP manifests topic properties (cf. Sect. 2.2), she proposes the following lexical entry for *bǎ* (p. 127):

(42) *bǎ* V (↑ PRED) = 'ba<SUBJ,OBJ,COMP>'
                (↑ OBJ) = (↑ COMP TOPIC)

According to (42), *bǎ* is a three-place verb, subcategorizing for a subject, an object, and a finite clause (i.e. the closed function COMP). The object corresponds to the post-*bǎ* NP, which simultaneously serves as the topic of the finite clause. The c- and f-structures for (43), given in (44), illustrate Bender's analysis.

(43) tā bǎ júzi bō-le pí ⟨M⟩
      3SG BA orange peel-PFV skin
      'S/he peeled the orange.'

(44)
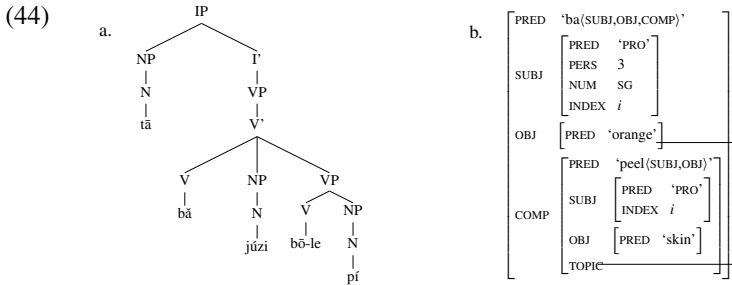
In (44b), (↑ OBJ) = (↑ COMP TOPIC) ensures that OBJ also functions as TOPIC in the embedded clause (i.e. COMP). This equating of two functions gives rise to structure sharing, notated by a line connecting TOPIC and the f-structure that is the value of OBJ. This formalism allows "the two functions to have the same f-structure as their value" (Falk 2001, p. 126).

By definition, the finite clause, i.e. COMP in (44b), should have its own subject, but in (43) and (44a) there is no overt subject within this clause. Consequently, we need to posit a covert subject for it to satisfy the *Completeness Condition*.

(45)　　Completeness Condition (Kaplan and Bresnan 1995, p. 65)
　　　　An f-structure is *locally complete* iff it contains all the governable grammatical functions that its predicate governs. An f-structure is *complete* iff it and all its subsidiary f-structures are locally complete.

This covert subject, as we can see in (44b), is an abstract pronoun (PRO) whose referent is determined by the context. In LFG terms, this PRO is *anaphorically controlled* by a contextually available NP (Dalrymple et al. 2019, Ch.15; Falk 2001, Sect. 5.2), as indicated by the INDEX attribute in (44b).[17]

Bender's (2000) analysis has several advantages over Her's (1990). On the one hand, the matrix subject is selected by *bǎ*, not the main verb, so it explains why sometimes the matrix subject is not related to the main verb (see (39)). On the other, it captures the post-*bǎ* NP's topic properties (Sect. 2.2), which Her (1990) ignores. Furthermore, Bender (2000, p. 137) cites Huang's (1992) observation that "object controllers (overt objects which control another function) cannot be structural [i.e. clause-initial] topics." This may be why the post-*bǎ* NP cannot be fronted, as it controls COMP TOPIC in (42).

However, Her (2009, pp. 451–453) points out that Bender's analysis leaves COMP SUBJ unaccounted for. In other words, the anaphoric control relation between COMP SUBJ and another function is not specified. Therefore, Her proposes to make the following revision:

(46)　　*bǎ*　V　(↑ PRED) = 'ba<SUBJ,OBJ,XCOMP>'
　　　　　　　　(↑ OBJ) = (↑ XCOMP TOPIC)
　　　　　　　　(↑ SUBJ) = (↑ XCOMP $\gamma(\hat{\theta})$)

---

[17] Coindexed phrases are coreferential. Representing INDEX in f-structure is a simplification. Dalrymple et al. (2019, pp. 572ff) offer a more comprehensive analysis in conformity with LFG's modular architecture.

Here, $\hat{\theta}$ represents the logical subject and γ is one of the correspondence functions in LFG's parallel architecture. $\gamma(\hat{\theta})$ maps the logical subject to its associated GF. Thus, the f-description $(\uparrow \text{SUBJ}) = (\uparrow \text{XCOMP } \gamma(\hat{\theta}))$ links *bǎ*'s subject with the logical subject of the first embedded clause. Her (2009, p. 447) also mentions in a footnote that replacing COMP with XCOMP is "more in line with the LFG conventions", presumably because the embedded clause is assumed to be an open function lacking a subject (Sect. 3.1).

To see how (46) works, let us first look at (44b), which has a retained object *pí* 'skin' in the embedded clause (COMP in (44b) but XCOMP as per (46)). The predicate of this embedded clause is *bō* 'peel', whose logical subject is its syntactic subject. Therefore, (46) successfully links the matrix subject *tā* 's/he' to the embedded subject. On the other hand, *bǎ*-sentences containing unmarked passives (Sect. 2.5) receive a different treatment:

(47) a. lǐsì bǎ nà jiān fángzi chāi-le   ⟨M⟩
          Lisi BA DEM CLF house demolish-PFV
          'Lisi demolished that house.'

   b.
   $$\begin{bmatrix} \text{PRED} & \text{'ba}\langle\text{SUBJ,OBJ,XCOMP}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Lisi'} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'house'} \end{bmatrix} \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{'demolish}\langle\text{SUBJ}\rangle\text{'} \\ \text{TOPIC} & \\ \text{SUBJ} & \\ \text{VOICE} & \text{PASSIVE} \end{bmatrix} \end{bmatrix}$$

In the f-structure, OBJ simultaneously serves as XCOMP TOPIC, as required by the second line of the f-description in (46). The third line of the f-description is not realized syntactically, because the logical subject of *chāi* 'demolish', i.e. Lisi, is demoted as a result of passivization. Although Her (2009) does not specify how XCOMP SUBJ is functionally controlled, it can be accounted for via the default lexical rule of functional control: $(\uparrow \text{OBJ}) = (\uparrow \text{XCOMP SUBJ})$ (Bresnan 1982a, p. 378).

A problem with Her's (2009) proposal concerns XCOMP, whose SUBJ is said to be "functionally or anaphorically controlled" (p. 451). This is incorrect because XCOMP, by definition, is the functionally controlled clause (Bresnan 1982a, p. 376; Dalrymple et al. 2019, Sect. 15.1). This is not just an issue of terminology, as empirical evidence shows the XCOMP assumption for the *bǎ/de*-construction is mistaken.

First, an XCOMP analysis predicts that split antecedents are impossible in *bǎ/de*-sentences (Bresnan 1982a, p. 396). However, the following examples show that 'old Zhang' and 'old Huang' can be split antecedents for the embedded subject.

(48) a. lǎo-zhāng$_i$ jiào lǎo-huáng$_j$ yìqǐ   bǎ shū [e_{i+j} sī-diào sān yè] ⟨M⟩
          old-Zhang ask old-Huang together BA book   rip-off three page
          'Old Zhang asked old Huang to rip three pages out of the book together.'

   b. lɛ^0-tɕɛ$_i^{44}$ tɕy^0 lɛ^0-ɦiɔ$_j^{31}$ ʔe^{0}tɕʰi^{35} de^0 səɯ^{44} [e_{i+j} do^{212} sɔ^{44} ji^{212} l-ɔ^{212}] ⟨W⟩
          old-Zhang ask old-Huang together DE book   rip   three page off-PFV
          'Old Zhang asked old Huang to rip three pages out of the book together.'

Second, the *de*-construction allows a retained subject (Sect. 2.6), e.g., *bei* 'peel' in (49), which would have been ruled out were the subject functionally identical to another phrase.

(49) gi^{31} de^0 $sei^{44}ku^{44}$ bei^{31} $t\varphi^hi^{323}$-d-$ɔ^0$ 〈W〉
     3SG DE watermelon peel eat-inside-PFV
     'S/he ate up the watermelon peel.'

Based on the discussion above, I propose the following lexical entry for *bǎ*:

(50) *bǎ* V (↑ PRED) = 'ba<SUBJ,OBJ,COMP>'
         (↑ OBJ) = (↑ COMP TOPIC)
         (↑ COMP SUBJ PRED) = 'PRO'
         @PASSIVE, @ACTIVE

The first two lines of the lexical entry are in line with Bender (2000): the verbal complement of *bǎ* is a COMP, whose TOPIC is identified with *bǎ*'s OBJ via structure sharing. The third line supplies a pronominal subject for the COMP to meet the Completeness Condition (45). The antecedent of this PRO is captured by the two templates (Dalrymple et al. 2004) defined in the fourth line. The templates serve to reformulate Her's (2009) insight that the referent of COMP SUBJ is predictable. The PASSIVE template is for *bǎ*-sentences with a passive main verb and the ACTIVE template is for *bǎ*-sentences with an active main verb, including those with a retained object and those that involve long-distance dependencies (see below). These templates are defined in (51):

(51) PASSIVE ≡ (↑ COMP VOICE) = PASSIVE ⇒
             (↑ OBJ INDEX) = (↑ COMP SUBJ INDEX)
     ACTIVE ≡ (↑ COMP VOICE) = ACTIVE ⇒
             (↑ SUBJ INDEX) = (↑ COMP SUBJ INDEX) ∧
             ((%EOBJ PRED) = 'PRO' ∧ (↑ OBJ INDEX) = (%EOBJ INDEX))
             (↑ COMP COMP$^+$ OBJ) = %EOBJ

Both templates contain an *if-then* clause, as indicated by the double right arrow.[18] Informally, the PASSIVE template states that "if my complement clause is in the passive voice, then my object is coindexed with the complement clause's subject." The first two lines of the ACTIVE template, which are sufficient to analyze retained objects, state that "if my complement clause is in the active voice, then my subject is coindexed with the complement clause's subject." The third line is optional, as indicated by the outermost brackets. It uses a *local name* %EOBJ (Kaplan and Maxwell III 1996, pp. 90–91), mnemonically for "embedded object", to abbreviate the constraint in the fourth line, (↑ COMP COMP$^+$ OBJ).[19] To put it in prose, this equation says "I can, but don't have to, introduce a pronominal object at some level of embedding, which must be at least two levels of embedding, and this embedded object is coreferential with my object."

---

[18] See Bresnan et al. (2016, p. 61) for the definition of the ⇒ conditional.

[19] The Kleene plus of a set of strings A, denoted A^+, refers to one or more occurrences of A. The Kleene star A^* refers to zero or more occurrences of A (Partee et al. 1990, Sect. 17.2).

To illustrate my proposal, consider the sentence and f-structure given in (52).[20]

(52)   a.   wǒ bǎ [zhè gè wèntí]$_i$ shèfǎ jiějué e_i ⟨M⟩
           1SG BA DEM CLF   problem attempt resolve
           'I attempted to resolve this problem.'

b.

$$
\begin{bmatrix}
\text{PRED} & \text{'ba}\langle\text{SUBJ,OBJ,COMP}\rangle\text{'} \\[4pt]
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PERS} & 1 \\ \text{NUM} & \text{SG} \\ \text{INDEX} & i \end{bmatrix} \\[20pt]
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'problem'} \\ \text{INDEX} & j \end{bmatrix} \\[12pt]
\text{COMP} & \begin{bmatrix}
\text{PRED} & \text{'attempt}\langle\text{SUBJ,COMP}\rangle\text{'} \\
\text{VOICE} & \text{ACTIVE} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{INDEX} & i \end{bmatrix} \\[10pt]
\text{COMP} & \begin{bmatrix}
\text{PRED} & \text{'resolve}\langle\text{SUBJ,OBJ}\rangle\text{'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{INDEX} & i \end{bmatrix} \\[10pt]
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{INDEX} & j \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

The post-*bǎ* verb *shèfǎ* 'attempt' is active. The lexical entry for *shèfǎ* includes 'attempt <SUBJ,COMP>', i.e. it selects a subject and a finite clause as its arguments. As a subject control verb, *shèfǎ* requires the embedded subject (the subject of *jiějué* 'resolve') to be coreferential with its local subject, but it says nothing about its embedded object, which is implicit in this case. Intuitively, this implicit object is coreferential with the post-*bǎ* NP, *zhè gè wèntí* 'this problem', so their relationship is long-distance.

In the f-structure in (52b), the matrix subject is indexed with i, and the matrix object j (the structure sharing between OBJ and COMP TOPIC is omitted for simplicity). Since *shèfǎ* 'attempt' is an active verb, the ACTIVE template in (50) is called for, which requires the matrix subject and the embedded subject to have the same index. Consequently, both SUBJ and COMP SUBJ in (52b) are indexed with i, i.e. they are coreferential. Within this COMP, *shèfǎ* 'attempt' introduces another level of embedding, so the optional equations in (51) are invoked: they assign the value 'PRO' to the object of *jiějué* 'resolve' to satisfy the Completeness Condition (45) and coindex this object with the post-*bǎ* NP.

Summarizing, this section has reviewed three LFG approaches to the *bǎ*-construction and their implications for the *de*-construction. The conclusions are (i) *bǎ* is a three-place verb, subcategorizing for a subject, an object, and a finite clause (Bender 2000); (ii) *bǎ*'s object is multi-functional as it also serves as the topic of the finite clause; and (iii) the subject of the finite clause is not functionally identical to *bǎ*'s subject or object, but may be coreferential with one of them in a predictable way (Her 2009). These observations are formalized in the lexical entry in (50).

---

[20] I thank an anonymous reviewer for suggesting this example.

Recall that the multi-functionality of the post-*bǎ* NP, i.e. simultaneously being OBJ and COMP TOPIC is the reason why this NP cannot be topicalized (Bender 2000). This means that the same analysis clearly cannot apply to both the *bǎ*-construction and the *de*-construction: Sect. 2.2 has presented evidence that the post-*de* NP has topic properties, so this NP is plausibly multi-functional as well (being both OBJ and COMP TOPIC). However, Sect. 2.1 has shown that the post-*de* NP can topicalize using the gap strategy, which is unexpected if the multi-functionality prohibits topicalization. This problem, as well as those disparities summarized in Sect. 2.7, will be addressed in the next section.

4 A formal analysis of *de*

Bender (2000) contends that *bǎ*'s object serves as the topic of the clausal complement (COMP TOPIC). The structural relation between the object and COMP TOPIC prohibits the object from taking the clause-initial topic position because, as Huang (1992) has shown, long-distance dependencies involving Mandarin object controllers are disallowed. If Bender's analysis is on the right track for Wenzhounese *de* as well, it implies that *de*'s object should not be equated with COMP TOPIC, because this object can be topicalized. By doing so, however, a crucial generalization is lost; that is, *de*'s object does have topic properties (Sect. 2.2). Therefore, a formal analysis of *de* needs to address this dilemma as well as capture the differences summarized in (33), repeated below as (53).

(53) a. In Mandarin, resumption is the only way to topicalize the post-*bǎ* NP. By contrast, Wenzhounese can use either resumption or the gap strategy.
      b. The resumptive pronoun in Mandarin shows (optional) number agreement, but its Wenzhounese counterpart GI does not.
      c. In the *bǎ*-construction, the quantifier uniformly follows the post-*bǎ* NP. In Wenzhounese, the quantifier follows the post-*de* NP when there is no topicalization; the quantifier precedes *de* when there is topicalization.
      d. There is a double *de*-construction which is unattested in Mandarin.
      e. Only Wenzhounese can contain a retained subject in the *de*-construction.

To capture these observations, I propose two lexical entries for *de*. De_1 is similar to but not the same as Mandarin *bǎ* and it prohibits extraction of the post-*de* NP. By contrast, *de*-sentences involving a post-*de* gap or a resumptive GI are exclusively related to de_2, whose object does not have topic properties *per se* and can therefore be topicalized.

In the *bǎ*-construction in Mandarin, the topic, if any, anaphorically binds the resumptive pronoun, while in Wenzhounese, the topic functionally binds the gap. The resumptive GI, as argued in Sect. 2, is a processor resumptive inserted after grammatical computation, so it need not be specified by the grammar (Asudeh 2012). This explains why GI does not inflect for number (53b) and why it is gap-like with regard to quantifier float (53c). Section 4.3 will argue that quantifier float is essentially a locality effect. For the double-*de* construction (53d), I will argue that the second *de* must be de_2. Mandarin does not have a de_2 equivalent, hence the absence of a double-*bǎ*

construction. Finally, if the clausal complement of *de* corresponds to IP in c-structure, it can license the retained subject (53e). By contrast, the clausal complement of *bǎ* needs to correspond to VP to avoid overgenerating a retained subject in Mandarin. The full analyses are presented below.

4.1 *De*$_1$

I postulate *de*$_1$ for *de*-sentences in which there is neither topicalization nor GI, so it is similar to the Mandarin *bǎ*. The only difference lies in the third line of (54).

(54)   *de*$_1$ V   ($\uparrow$ PRED) = 'de<SUBJ,OBJ,COMP>'
                        ($\uparrow$ OBJ) = ($\uparrow$ COMP TOPIC)
                        (($\uparrow$ COMP SUBJ PRED) = 'PRO')
                        @PASSIVE, @ACTIVE

Just as *bǎ* does in (50), so *de*$_1$ subcategorizes for a subject, an object, and a finite clause. The second line in (54) equates OBJ with COMP TOPIC, which explains the topic properties of this OBJ (Sect. 2.2) and prohibits it from undergoing topicalization. When the subject of the COMP is implicit, the f-description ($\uparrow$ COMP SUBJ PRED) = 'PRO' provides a covert subject to satisfy the Completeness Condition (45). This is where *de*$_1$ differs from *bǎ*: this equation is optional (indicated by the outermost brackets; see Dalrymple et al. 2019, Sect. 5.2.4 for optionality) for *de*$_1$ but obligatory for *bǎ*. Optionality is necessarily for *de*$_1$ because a retained subject (Sect. 2.6) blocks the application of this equation, for the sake of *Coherence*.

(55)   Coherence Condition (Kaplan and Bresnan 1995, p. 65)
          An f-structure is *locally coherent* iff all the governable grammatical functions that it contains are governed by a local predicate. An f-structure is *coherent* iff it and all its subsidiary f-structures are locally coherent.

The phrase structure rules relevant for *de*$_1$ are listed in (56). Note that the subject NP in (56a) is optional because Wenzhounese is a discourse pro-drop language.

(56)   a.   IP $\rightarrow$ $\begin{pmatrix} \text{NP} \\ (\uparrow \text{ SUBJ}) = \downarrow \end{pmatrix}$ I'
                                                $\uparrow = \downarrow$

         b.   V' $\rightarrow$ V   NP   IP
                              $\uparrow = \downarrow$ ($\uparrow$ OBJ) = $\downarrow$ ($\uparrow$ COMP) = $\downarrow$

My proposal successfully generates the c- and f-structures for simple *de*-sentences, *de*-sentences with a retained subject or a retained object, and those which involve long-distance dependencies. First, (57) gives a simple *de*-sentence.

(57)   lɛ^0-tɕɛ^{44} de^0 pɛ^{0}ɕi^{53}-ɕɛ^{44} ta^{35}kʰe^{44} ⟨W⟩
          old-Zhang DE secure-box open
          'Old Zhang opened the safe.'

(58)
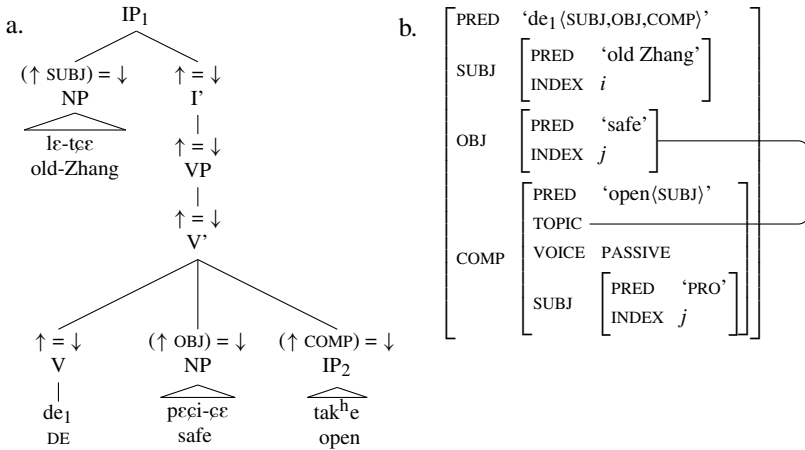
In (58a), V' immediately dominates V', NP, and IP$_2$, the latter two corresponding to f-structural OBJ and COMP, respectively. This flat structure suffices for the present purpose, as (I) there is no solid evidence that [V NP] or [NP IP$_2$] form a constituent (cf. Feng 2019; Huang et al. 2009), and none of my analysis hinges on the flat structure.

In (58b), the matrix object simultaneously functions as COMP TOPIC, as required by the second line of de_1's lexical entry in (54).[21] The embedded subject (COMP SUBJ) is a covert pronoun, supplied by the third line of (54). Since the embedded verb is an unmarked passive (Sect. 2.5), it invokes the PASSIVE template of de's lexical entry, so both de's object and COMP SUBJ are indexed as j in the f-structure.

The second example (59) has a retained subject. This motivates the IP analysis of de's clausal complement, as subjects are placed in Spec,IP in Wenzhounese. By contrast, a VP analysis is sufficient for $bǎ$'s clausal complement (Bender 2000) because the $bǎ$-construction does not license a retained subject.

(59)　lɛ^0-tɕɛ^{44}　de^0 sei^{44}ku^{44}　bei^{31} tɕʰi^{323}-d-ɔ^{22}　⟨W⟩
　　　old-Zhang DE watermelon peel　eat-inside-PFV
　　　'Old Zhang ate up the watermelon peel.'

In (60a), bei 'peel' occupies Spec,IP and maps to the embedded subject (COMP SUBJ) in (60b). This overt COMP SUBJ blocks the application of ($\uparrow$ COMP SUBJ PRED) = 'PRO' in (54) to satisfy Coherence (55), which is possible because the equation is optional. The whole-part relationship between $seiku$ 'watermelon' and bei 'peel' is not represented in (59b) for ease of exposition, but it could be easily incorporated by positing a POSS(essor) attribute for COMP SUBJ and having it be functionally bound by COMP TOPIC.

---

[21] In what follows, I will omit COMP TOPIC if it has little bearing on the discussion.

(60)   a.

$$\ldots$$
$$\text{V'}$$

↑ = ↓	(↑ OBJ) = ↓	(↑ COMP) = ↓
V	NP	IP
de$_1$	seiku	
DE	watermelon	

	(↑ SUBJ) = ↓	↑ = ↓
	NP	I'
	bei	↑ = ↓
	peel	VP

tɕʰi-d-ɔ
eat-inside-PFV

b.
$$
\begin{bmatrix}
\text{PRED} & \text{'de}_1 \langle \text{SUBJ,OBJ,COMP} \rangle \text{'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'old Zhang'} \\ \text{INDEX} & i \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'watermelon'} \\ \text{INDEX} & j \end{bmatrix} \\
\text{COMP} & \begin{bmatrix} \text{PRED} & \text{'eat-inside} \langle \text{SUBJ} \rangle \text{'} \\ \text{VOICE} & \text{PASSIVE} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'peel'} \end{bmatrix} \end{bmatrix}
\end{bmatrix}
$$
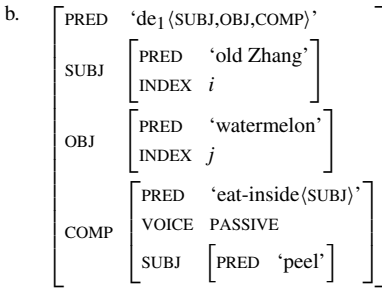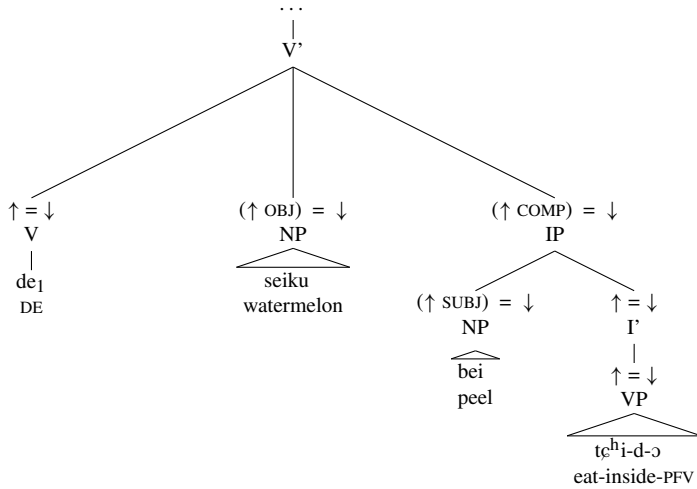
The third example (61) contains a retained object. The c-structure is unremarkable for the analysis, so only its f-structure is given in (62).

(61)   lɛ^0-tɕɛ^{44} de^0 sei^{44}ku^{44} tɕʰi^{323} ʔe^{0}pø^{53} d-ɔ^{22} ⟨W⟩
old-Zhang DE watermelon eat   half   inside-PFV
'Old Zhang ate up half of the watermelon.'

(62)
$$
\begin{bmatrix}
\text{PRED} & \text{'de}_1 \langle \text{SUBJ,OBJ,COMP} \rangle \text{'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'old Zhang'} \\ \text{INDEX} & i \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'watermelon'} \\ \text{INDEX} & j \end{bmatrix} \\
\text{COMP} & \begin{bmatrix} \text{PRED} & \text{'eat} \langle \text{SUBJ,OBJ} \rangle \text{'} \\ \text{VOICE} & \text{ACTIVE} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{INDEX} & i \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'half'} \end{bmatrix} \end{bmatrix}
\end{bmatrix}
$$
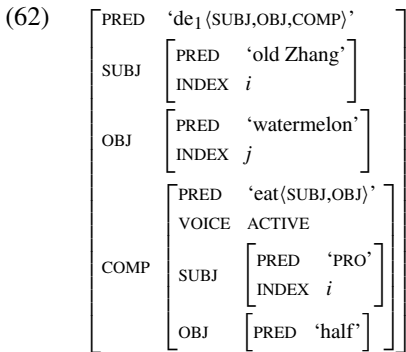
Here, the embedded verb $t\varepsilon^{h}i$ 'eat' is active, and it takes an object $\textipa{P}ep\phi$ 'half'. The ACTIVE template requires the matrix subject to anaphorically control the embedded subject, so SUBJ and COMP SUBJ share the index i.

Finally, the example in (63) involves a long-distance dependency between *de*'s object and the object of $tak^{h}e$ 'open'.

(63) $l\varepsilon^{0}$-$t\varepsilon\varepsilon^{44}$ de^{0} $p\varepsilon^{0}\varepsilon i^{53}$-$\varepsilon\varepsilon_{i}^{44}$ [$\varepsilon\varepsilon^{0}$ $b\mathfrak{o}^{13}f\mathfrak{o}^{323}$ [$ta^{35}k^{h}e^{44}$ e_{i}]] $\langle W \rangle$
       old-Zhang DE secure-box   think method   open
       'Old Zhang thought of a way to open the safe.'

(64)

$$
\begin{bmatrix}
\text{PRED} & \text{'de}_1\langle\text{SUBJ,OBJ,COMP}\rangle\text{'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'old Zhang'} \\ \text{INDEX} & i \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'safe'} \\ \text{INDEX} & j \end{bmatrix} \\
\text{COMP} & \begin{bmatrix}
\text{PRED} & \text{'think}\langle\text{SUBJ,OBJ,COMP}\rangle\text{'} \\
\text{VOICE} & \text{ACTIVE} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{INDEX} & i \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'method'} \end{bmatrix} \\
\text{COMP} & \begin{bmatrix}
\text{PRED} & \text{'open}\langle\text{SUBJ,OBJ}\rangle\text{'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{INDEX} & i \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{INDEX} & j \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

In the f-structure in (64), all the subject functions are indexed as i. This anaphoric control relation is established as follows. First, the post-*de* verb, $\varepsilon\varepsilon$ 'think', does not have an overt subject, so the optional equation in (54) needs to apply to provide a covert subject for 'think'. Second, because 'think' is an active verb, it calls for the ACTIVE template in de_1's lexical entry. According to the template's definition in (51), the subject of *de* is coindexed with the subject of 'think'. Third, this verb (more precisely, the idiomatic expression $\varepsilon\varepsilon$ $b\mathfrak{o}f\mathfrak{o}$ 'think of a way') is a subject-control verb, so its lexical entry carries the information to coindex its subject with the subject of its clausal complement. In this case, COMP SUBJ and COMP COMP SUBJ both bear the index i.

On the other hand, *de*'s object anaphorically controls the most embedded object. Their coindexation is also achieved by the ACTIVE template in (51). The last two lines of the template explicitly refer to COMP COMP^{+} OBJ, which is the object of $tak^{h}e$ 'open' in this case. The equation (%EOBJ PRED) = 'PRO' supplies a predicate value to this covert object, while the equation ($\uparrow$ OBJ INDEX) = (%EOBJ INDEX) mandates anaphoric control between the object of *de* and the object of 'open'.

In short, the lexical entry for de_1 (54) is similar to that for *bǎ* (50), and the only difference is that *bǎ* obligatorily supplies an implicit subject for its embedded clause whereas

de$_1$ does so optionally. This is motivated by the retained subject in *de*-sentences, which excludes an implicit subject due to the Coherence Condition. The proposed lexical entry and phrase structure rules successfully captures *de*-sentences of variant complexities (retained subjects, retained objects, and long-distance dependencies). Nevertheless, the data involving topicalization are still missing from the analysis. I will discuss these in the next section.

4.2 *De*$_2$

To address the topicalization paradox, I contend that there are two different elements *de* in Wenzhounese. *De*$_1$, discussed in the previous section, is the one whose object has topic properties and therefore cannot be topicalized. By contrast, *de*$_2$'s object does not have topic properties. Instead, its object is essentially a gap that must be identified with a topic. The two lexical entries for *de*, though distinct, are also related, which can be formalized as a lexical redundancy rule (Bresnan 1982b; Jackendoff and Audring 2020). There may also be a diachronic relation between *de*$_1$ and *de*$_2$, which I leave for future research.

As we shall see, proposing two lexical entries is not just an *ad hoc* way to dodge the topicalization paradox; rather, it has empirical advantages in accounting for the distribution of the quantifier and the double-*de* construction.

The lexical entry for *de*$_2$ is given in (65a). The first three lines are already familiar to us: they state that *de*$_2$ is a three-place verb and it optionally provides a pronominal subject to its clausal complement. It also uses the templates defined in (51) to resolve the anaphoric control relation between *de*$_2$'s local GFs and more embedded GFs.

(65)   *de*$_2$ V   ($\uparrow$ PRED) = 'de<SUBJ,OBJ,COMP>'
                        (($\uparrow$ COMP SUBJ PRED) = 'PRO')
                        @PASSIVE, @ACTIVE
                        $\phi^{-1}(\uparrow$ OBJ) $= \emptyset$
                        ($\uparrow$ OBJ) = ((COMP* $\uparrow$) TOPIC)

The fourth line begins with ϕ^{-1}, a correspondence relation that maps an f-structure to a c-structural node (Dalrymple et al. 2019, Ch.4). Informally, it says "my object does not correspond to any c-structural node." Note that this does not contradict the presence of the processor resumptive GI, since GI is licensed not by the grammar, but by the post-grammatical processing model.

The fifth line contains an operation on f-structure called *inside-out functional uncertainty*, notated (COMP* $\uparrow$).[22] Informally, the equation ($\uparrow$ OBJ) = ((COMP* $\uparrow$) TOPIC) can be thought of as a feature path exiting from the f-structure immediately containing OBJ to superordinate f-structures, with the only feature along this path being COMP. This feature path must terminate at a TOPIC attribute in a local or superordinate f-structure. Take (66) for example, the start point is the f-structure labelled *c*, from which there is a path to *b* along the COMP attribute in *b*. The path does not lead to *a* because there is

---

[22] The Kleene star is defined in footnote 19. The formal definition of the inside-out expression is as follows: $(af) = g$ holds iff g is an f-structure, a is a symbol, and the pair $\langle a, f \rangle \in g$ (Dalrymple et al. 2019, p. 210).

no COMP in *a*. Given that there is a TOPIC in *b*, the feature path successfully terminates, so it identifies (*b* TOPIC) with (*c* OBJ).

(66)

$$
a\begin{bmatrix} \dots & \dots \\ \text{XCOMP} & b\begin{bmatrix} \text{TOPIC} & \dots \\ \text{COMP} & c\begin{bmatrix} \dots & \dots \\ \text{OBJ} & \dots \end{bmatrix} \end{bmatrix} \end{bmatrix}
$$

The phrase structure rules relevant for constructions that include *de$_2$* are listed in (67). The rule in (67a) is a general rule for topicalization. Topics are adjoined to IP, following Bresnan et al. (2016). Since (at least theoretically) there could be more than one topic, the relevant rule includes a set membership symbol rather than an equals sign: $\downarrow \in$ (↑ TOPIC) instead of (↑ TOPIC) = $\downarrow$. This means that the value of TOPIC is a set rather than a single f-structure. The topic phrase is also annotated with (↑ COMP* GF) = $\downarrow$,[23] so it will bear some grammatical function (GF) within the clause, as required by the Extended Coherence Condition in (68).

(67)   a.   IP   →   NP   IP
                           $\downarrow \in$ (↑ TOPIC)   ↑=$\downarrow$
                          ((↑ COMP* GF) = $\downarrow$)
         b.   V'   →   V   IP
                        ↑=$\downarrow$ (↑ COMP) = $\downarrow$

(68)   Extended Coherence Condition (Dalrymple et al. 2019, p. 653)
         FOCUS and TOPIC must be linked to the semantic predicate argument structure
         of the sentence in which they occur, either by functionally or by anaphorically
         binding an argument.

I will illustrate my proposal with (69), in which the post-*de* object is a gap. Note that the grammar that generates (69) is the same as the grammar that generates a corresponding sentence with the resumptive GI, because the insertion of GI is manipulated by performance, not competence.

(69)   ge^{212}-le^0 sei^{44}ku$_i^{44}$,   lɛ^0-tɕɛ^{44} de^0 Ø$_i$ tɕʰi^{323}-d-ɔ^{22} ⟨W⟩
         DEM- PL watermelon old-Zhang DE   eat-inside-PFV
         'These watermelons, old Zhang ate them.'

---

[23] This annotation is only an approximation, as functions other than COMP may be allowed in the feature path. The precise formulation is orthogonal to this paper, and the reader is referred to Huang (1992, 1993) for discussions of long-distance dependency in Mandarin in the LFG setting.

(70) a.

IP
- $\downarrow \in (\uparrow \text{TOPIC})$
  NP
  ge-le seiku
  these watermelons
- $\uparrow = \downarrow$
  IP
  - $(\uparrow \text{SUBJ}) = \downarrow$
    NP
    lɛtɕɛ
    old Zhang
  - $\uparrow = \downarrow$
    I'
    - $\uparrow = \downarrow$
      VP
      - $\uparrow = \downarrow$
        V'
        - $\uparrow = \downarrow$
          V
          de
          DE
        - $(\uparrow \text{COMP}) = \downarrow$
          IP
          tɕʰi d-ɔ
          eat inside-PFV

b.

$$
\begin{bmatrix}
\text{PRED} & \text{`de}_2\langle\text{SUBJ,OBJ,COMP}\rangle\text{'} \\
\text{TOPIC} & \left\{\begin{bmatrix} \text{PRED} & \text{`watermelon'} \\ \text{NUM} & \text{PL} \\ \text{DEIXIS} & \text{PROX} \\ \text{INDEX} & i \end{bmatrix}\right\} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`old Zhang'} \end{bmatrix} \\
\text{OBJ} & \\
\text{COMP} & \begin{bmatrix} \text{PRED} & \text{`eat}\langle\text{SUBJ}\rangle\text{'} \\ \text{VOICE} & \text{PASSIVE} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`PRO'} \\ \text{INDEX} & i \end{bmatrix} \end{bmatrix}
\end{bmatrix}
$$
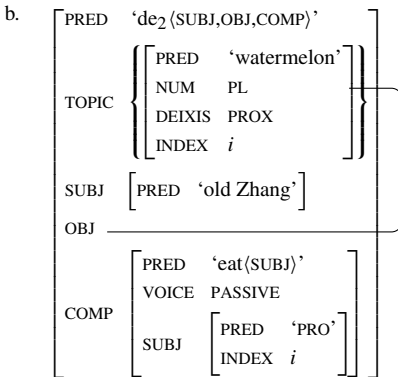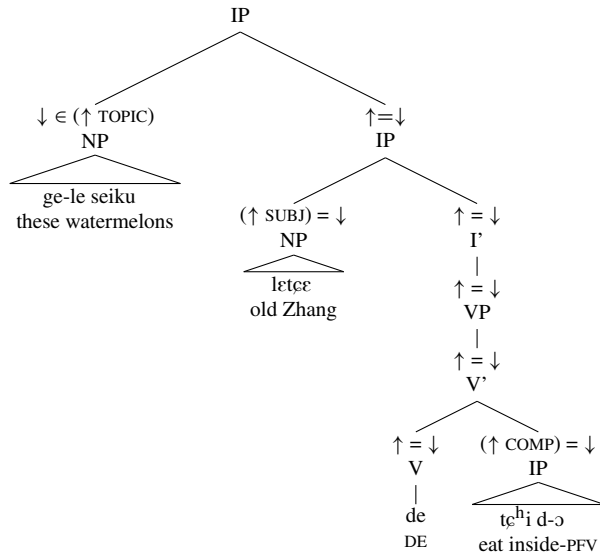
In (70a), V' does not dominate a locally realized NP, in conformity to the f-description $\phi^{-1}(\uparrow \text{OBJ}) = \emptyset$ in de_2's lexical entry (65). In (70b), TOPIC is equated with OBJ, as indicated by the curved line. This structure sharing relation can be established either by the general rule for topicalization (67a) or the inside-out expression in (65), but this does not mean that the latter is redundant, as we shall see in the analysis of the double-*de* construction.

The analyses of retained subjects, retained objects, and long-distance dependencies are similar to those of de_1, the difference being that de_2's object does not correspond to a c-structure node and must be equated with a topic. The remaining sections will demonstrate how my proposals above account for the distribution of the quantifier and the double-*de* construction.

4.3 Quantifier float and locality

Much work on the Mandarin universal quantifier has been done in the Government and Binding framework (see Cheng 1995; Chiu 1993, and references therein). Cheng (1995) observes that the universal quantifier *dōu* needs to be local to its NP, which is *xuéshēng* 'student' in (71a), or the trace of *xuéshēng* in (71b).

(71)   a.   zhè-xiē   xuéshēng (<u>dōu</u>) zhīdào [wǒ (*dōu) xǐhuān biàn-zhuāng]   ⟨M⟩
          DEM- PL student   (all)   know 1SG (*all)   like   change-costume
          'All these students know that I like to be in drag.'

     b.   [zhè-xiē xuéshēng]$_i$, wǒ   (*dōu) xiāngxìn [lǐsì (<u>dōu</u>) hěn   xǐhuān $\emptyset_i$] ⟨M⟩
          DEM- PL student   1SG (*all)   believe   Lisi   (all)   very like
          'These students, I believe that Lisi likes them all.'

As for the universal quantifier in Wenzhounese, Sect. 2.3 has shown that it is usually adjacent to the post-*de* NP (72a), but when the post-*de* NP is topicalized, the quantifier tends to raise to the higher clause where the topic is (72b).

(72)   a.   ŋ^{13} de^0 lɔ^{25}dʑa^{212} <u>ʔo^{323}</u> tɛ^{53}-k^{h}ɔ^0 ⟨W⟩
          1SG DE trash   all   dump-away
          'I dumped all of the trash.'

     b.   lɔ^{25}dʑa$_i^{212}$, ŋ^{13} <u>ʔo^{323}</u> de^0 $\emptyset_i$ tɛ^{53}-k^{h}ɔ^0 ⟨W⟩
          trash   1SG all   DE   dump-away
          'The trash, I dumped it all.'

    c.??lɔ^{25}dʑa$_i^{212}$, ŋ^{13} de^0 $\emptyset_i$ <u>ʔo^{323}</u> tɛ^{53}-k^{h}ɔ^0 ⟨W⟩
          trash   1SG DE   all   dump-away
          'The trash, I dumped it all.'

(72a) is similar to (71a), in which the quantifiers are local to their NP. The disfavoured but still grammatical example in (72c) is similar to (71b), where the quantifiers are local to the trace of the NP. However, (72b) contrasts with the Mandarin examples in that the quantifier is local to the topicalized NP instead of its trace.

Therefore, it seems that locality is also relevant to quantifier float in the *de*-construction, although it differs from the locality condition for Mandarin *dōu*. The exact formulation of this locality condition is beyond the scope of this paper, due to the following pending issues. First, given that the quantifier is adjoined to the verbal domain in c-structure, what annotations are required to map it to the nominal domain in f-structure to model its scope? A typical annotation like (↓ ADJ) ∈ ↑ would wrongly predict that the quantifier modifies the verb. Second, an embedded quantifier as in (72c) is still grammatical. How do we capture this optionality and gradient grammaticality?

Despite these pending issues, the generalization we can make from the data above still corroborates the proposed analyses for *de*. Specifically, (71) suggests that the locality condition in Mandarin holds between the quantifier and the complement of the verb, be it in-situ or ex-situ (Cheng 1995). In LFG terms, the quantifier is local to OBJ at f-structure. In the *de*-construction (72), the locality condition holds between the

quantifier and the discourse function TOPIC, instead of OBJ.[24] In (72a), the quantifier is embedded in the clause (COMP) headed by $t\varepsilon$-$k^h\jmath$ 'dump-away', where there is a TOPIC bound by *lɔdʑa* 'trash', so the locality condition is met. In (72b), the matrix verb is de_2, so its object is not equated with a TOPIC according to (65). Consequently, the quantifier floats to the matrix clause to satisfy the locality condition with the matrix TOPIC.

In brief, assuming that the universal quantifier needs to be local to the *de*-sentence topic, its distribution naturally follows from the lexical entries for *de*. De_1's complement clause contains a topic, so the quantifier is placed within this clause. De_2's complement clause does not contain a topic, so the quantifier is placed in the matrix clause where there is a clause-initial topic. The processor resumptive GI does not block quantifier float because it is not part of the grammatical computation (Asudeh 2012).

4.4 The double-*de* construction

Section 2.4 discussed the double-*de* construction in Wenzhounese, which has no equivalent in Mandarin. This construction is constrained in the sense that the second *de* must be followed by a gap or the processor resumptive GI, as summarized in (73).

(73)   a.   Subject **de** Object **de** (GI) V
       b.   *Subject **de** Object **de** Object V
       c.   *Topic$_i$ Subject **de** Ø$_i$ **de** (GI) V
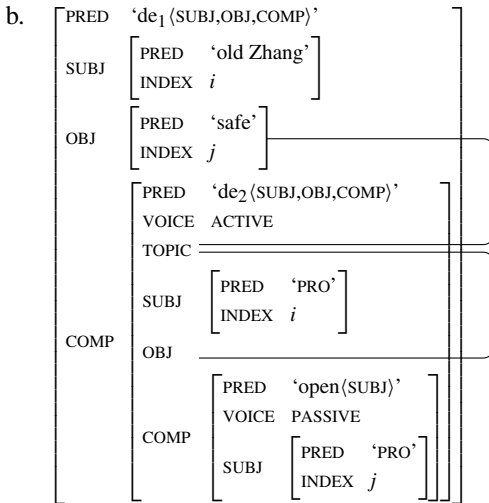   d.   Topic$_i$ Subject **de** Ø$_i$ V$_1$ **de** (GI) V$_2$

The generalization for this constraint directly follows from the distinct lexical entries for *de*: while both de_1 and de_2 can be the first *de* in the double-*de* construction,[25] only de_2 can be the second *de*. If we assume that Mandarin only has one lexical entry for *bǎ* (which is similar to de_1), the absence of a double-*bǎ* construction is unsurprising: there is no de_2 counterpart to license a second *bǎ*.

A further question is why the second *de* must be de_2. Recall the phrase structure rules in (56) and (67) readily allow the generation of the double-*de* construction, but they do not impose an order between de_1 and de_2. The explanation lies in the f-structure because the (un)grammaticality of the double-*de* construction relates to the arguments of *de*, not its constituency. Consider first a legitimate double-*de* sentence in (74).

(74)   a. $l\varepsilon^0$-$t\varepsilon\varepsilon^{44}$ de_1^0 $p\varepsilon^0\varepsilon i^{53}$-$\varepsilon\varepsilon_i^{44}$ de_2^0 Ø$_i$ $ta^{35}k^he^{44}$ ⟨W⟩
             old-Zhang DE secure-box DE   open
             'Old Zhang opened the safe.'

---

[24] That the topic can launch a floating quantifier is also observed in languages like Hebrew (Spector 2009) and Tagalog (Schachter 1976, p. 500ff).

[25] ((73c)) is possibly ruled out by a phonological constraint against adjacent instances of *de*.

b.
$$
\begin{bmatrix}
\text{PRED} & \text{'de}_1\,\langle\text{SUBJ,OBJ,COMP}\rangle\text{'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'old Zhang'} \\ \text{INDEX} & i \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'safe'} \\ \text{INDEX} & j \end{bmatrix} \\
\text{COMP} & \begin{bmatrix}
\text{PRED} & \text{'de}_2\,\langle\text{SUBJ,OBJ,COMP}\rangle\text{'} \\
\text{VOICE} & \text{ACTIVE} \\
\text{TOPIC} & \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{INDEX} & i \end{bmatrix} \\
\text{OBJ} & \\
\text{COMP} & \begin{bmatrix}
\text{PRED} & \text{'open}\langle\text{SUBJ}\rangle\text{'} \\
\text{VOICE} & \text{PASSIVE} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{INDEX} & j \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

The matrix predicate is de_1, whose clausal complement is headed by de_2. Assuming that both instances of *de* are active verbs,[26] the ACTIVE template defined in (51) will require the matrix SUBJ to anaphorically control COMP SUBJ, as indicated by their coindexation. The line connecting matrix OBJ and COMP TOPIC indicates the structure sharing relation stipulated in de_1's lexical entry (54). By contrast, de_2's lexical entry (65) requires its object to be equated with a TOPIC at a local or higher f-structure. Such a TOPIC is locally available in (74b). Also note that this TOPIC is not a clause-initial topic licensed by the phrase structure rule (67a), which justifies the necessity of the inside-out f-description in de_2's lexical entry (65). Finally, because tak^he 'open' is an unmarked passive, the PASSIVE template in de_2's lexical entry coindexes de_2's object, i.e. COMP OBJ in (74b), with COMP COMP SUBJ.

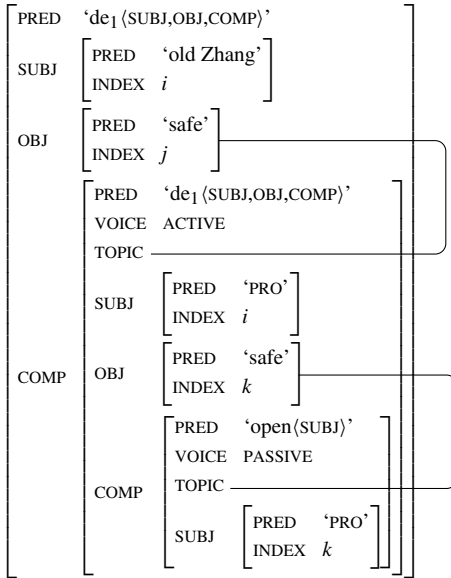An ungrammatical double-*de* sentence is given in (75a), in which both tokens of *de* take a full NP object.

(75)   a. $*l\varepsilon^0\text{-}t\varepsilon\varepsilon^{44}$ $de_1^0\ p\varepsilon^0\varepsilon i^{53}\text{-}\varepsilon\varepsilon^{44}$ $de_1^0\ p\varepsilon^0\varepsilon i^{53}\text{-}\varepsilon\varepsilon^{44}$ $ta^{35}k^he^{44}$ ⟨W⟩
           old-Zhang DE secure-box   DE secure-box   open
           Intended: 'Old Zhang opened the safe.'

---

[26] This is because *de* can never be passivized, similar to Mandarin *bǎ*. According to Her (2009), the *bǎ*-construction is the active counterpart of the passive *bei*-construction.

b. Ill-formed f-structure:

$$
\begin{bmatrix}
\text{PRED} & \text{`de}_1\,\langle\text{SUBJ,OBJ,COMP}\rangle\text{'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`old Zhang'} \\ \text{INDEX} & i \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`safe'} \\ \text{INDEX} & j \end{bmatrix} \\
\text{COMP} & \begin{bmatrix}
\text{PRED} & \text{`de}_1\,\langle\text{SUBJ,OBJ,COMP}\rangle\text{'} \\
\text{VOICE} & \text{ACTIVE} \\
\text{TOPIC} & \text{---} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`PRO'} \\ \text{INDEX} & i \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`safe'} \\ \text{INDEX} & k \end{bmatrix} \\
\text{COMP} & \begin{bmatrix}
\text{PRED} & \text{`open}\langle\text{SUBJ}\rangle\text{'} \\
\text{VOICE} & \text{PASSIVE} \\
\text{TOPIC} & \text{---} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`PRO'} \\ \text{INDEX} & k \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

In (75b), the matrix OBJ also functions as COMP TOPIC. Usually, this COMP TOPIC needs to anaphorically bind an implicit pronoun, a retained object, or a retained subject (Sect. 4.1), or it needs to share its structure with *de*$_2$'s OBJ, as in (74b). However, in (75b), no such relation can be established. This is because the value of a PRED attribute is a semantic form that is unique: each instance of use of the word *safe* gives rise to a uniquely instantiated occurrence of the semantic form 'safe' (Dalrymple et al. 2019, p. 45). Therefore, the two occurrences of 'safe' in (75b) must bear distinct indexes, which rules out the possibility of anaphoric control. As such, the relationship between COMP TOPIC and its local f-structure is unresolved, resulting in the ungrammaticality.

5 Conclusion

In this paper, I have investigated the syntax of the *de*-construction in Wenzhounese. Section 2 compared *de* with its Mandarin counterpart *bǎ* and revealed six differences as summarized in (53). In Sect. 3, I introduced LFG and critically reviewed three LFG approaches to the *bǎ*-construction. It was shown that *bǎ* and *de* should be analyzed as verbs with their own argument structures. Moreover, given the differences between *bǎ* and *de*, none of these analyses was directly applicable to the *de*-construction. Therefore, Sect. 4 proposed the following lexical entries for *de* and demonstrated that they could explain all the differences outlined in (53).

(76)   Lexical entries for *de*

   a. de_1 V $(\uparrow$ PRED$) =$ 'de$<$SUBJ,OBJ,COMP$>$'
                  $(\uparrow$ OBJ$) = (\uparrow$ COMP TOPIC$)$
                  $((\uparrow$ COMP SUBJ PRED$) =$ 'PRO'$)$
                  @PASSIVE, @ACTIVE

   b. de_2 V $(\uparrow$ PRED$) =$ 'de$<$SUBJ,OBJ,COMP$>$'
                  $((\uparrow$ COMP SUBJ PRED$) =$ 'PRO'$)$
                  @PASSIVE, @ACTIVE
                  $\phi^{-1}(\uparrow$ OBJ$) = \emptyset$
                  $(\uparrow$ OBJ$) = ((\text{COMP*}\ \uparrow)\ $ TOPIC$)$

(77)   Definitions of templates

   a.   PASSIVE $\equiv$ $(\uparrow$ COMP VOICE$) =$ PASSIVE $\Rightarrow$
                  $(\uparrow$ OBJ INDEX$) = (\uparrow$ COMP SUBJ INDEX$)$

   b.   ACTIVE $\equiv$ $(\uparrow$ COMP VOICE$) =$ ACTIVE $\Rightarrow$
                  $(\uparrow$ SUBJ INDEX$) = (\uparrow$ COMP SUBJ INDEX$) \wedge$
                  $((\%$EOBJ PRED$) =$ 'PRO' $\wedge (\uparrow$ OBJ INDEX$) = (\%$EOBJ INDEX$))$
                  $(\uparrow$ COMP COMP$^+$ OBJ$) = \%$EOBJ

Overall, this paper presents novel data from Wenzhounese, which not only are worthy of investigation in their own right, but may also shed new light on the study of *bǎ*. My study also contributes typologically and thematically to the research topics of Generative Grammar in general and LFG in particular.

Declarations

Conflict of interest The author declares that there is no conflict of interest.

References

Asudeh, Ash. 2011. Towards a unified theory of resumption. In *Resumptive pronouns at the interfaces*, ed. Alain Rouveret, 121–187. Amsterdam: John Benjamins.

Asudeh, Ash. 2012. *The logic of pronominal resumption*. Oxford: Oxford University Press.

Bao, Zhiming. 2011. Chinese tone sandhi. In *The Blackwell companion to phonology*, eds. Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, 1–25. Oxford: Wiley Online Library. https://doi.org/10.1002/9781444335262.wbctp0107.

Bender, Emily. 2000. The syntax of Mandarin bǎ: Reconsidering the verbal analysis. *Journal of East Asian Linguistics* 9: 105–145.

Bögel, Tina. 2015. The syntax-prosody interface in Lexical Functional Grammar. Ph.D. Thesis, Universität Konstanz.

Börjars, Kersti, Christopher Hicks, and John Payne. 2018. Interdependencies in Chinese noun phrases. In *Proceedings of the LFG18 conference*, ed. Miriam Butt and Tracy Holloway King, 109–128. Stanford: CSLI Publications.

Börjars, Kersti, Rachel Nordlinger, and Louisa Sadler. 2019. *Lexical-functional grammar: An introduction*. Cambridge: Cambridge University Press.

Bresnan, Joan. 1982a. Control and complementation. *Linguistic Inquiry* 13 (3): 343–434.

Bresnan, Joan. 1982b. The passive in lexical theory. In *The mental representation of grammatical relations*, ed. Joan Bresnan, 3–86. Cambridge: The MIT press.

Bresnan, Joan, Ash Asudeh, Ida Toivonen, and Stephen Wechsler. 2016. *Lexical-functional syntax*, 2nd ed. West Sussex: Wiley Blackwell.

Bresnan, Joan, and Sam A. Mchombo. 1995. The lexical integrity principle: Evidence from Bantu. *Natural Language & Linguistic Theory* 13: 181–254.

Carnie, Andrew. 2013. *Syntax: A generative introduction*, 2nd ed. West Sussex: Wiley Blackwell.

Chao, Yuen-Ren. 1930. A system of tone letters. *Le maître phonétique* 45: 24–27.

Chen, Matthew Y. 2000. *Tone sandhi: Patterns across Chinese dialects*. Cambridge: Cambridge University Press.

Chen, Yina. 2010. 瓯语中的"把"字句 [The de-construction in Wenzhounese]. *Journal of Language and Literature Studies* 7: 21–38.

Cheng, Lisa Lai-Shen. 1995. On Dou-quantification. *Journal of East Asian Linguistics* 4: 197–234.

Chiu, Bonnie. 1993. The inflectional structure of Mandarin Chinese. Ph.D. Thesis, University of California, Los Angeles.

Coltheart, Max. 1999. Modularity and cognition. *Trends in Cognitive Sciences* 3 (3): 115–120.

Dalrymple, Mary, Ronald M Kaplan, and Tracy Holloway King. 2004. Linguistic generalizations over descriptions. In *Proceedings of the LFG04 conference*, ed. Miriam Butt and Tracy Holloway King, 199–208. Stanford: CSLI Publications.

Dalrymple, Mary, John J. Lowe, and Louise Mycock. 2019. *The Oxford reference guide to Lexical Functional Grammar*. Oxford: Oxford University Press.

Engdahl, Elisabet. 1982. Restrictions on unbounded dependencies in Swedish. In *Readings on unbounded dependencies in Scandinavian languages*, ed. Elisabet Engdahl and Eva Ejerhed, 151–174. Stockholm: Almquist / Wiksell International.

Falk, Yehuda N. 2001. *Lexical-functional grammar: An introduction to parallel constraint-based syntax*. Stanford: CSLI Publications.

Feng, Shengli. 2019. *Prosodic syntax in Chinese: History and changes*. Abingdon: Routledge.

Ferreira, Fernanda, and Benjamin Swets. 2005. The production and comprehension of resumptive pronouns in relative clause 'island' contexts. In *Twenty-first century psycholinguistics: Four cornerstones*, ed. Anne Cutler, 263–278. Mahway: Lawrence Erlbaum Associates.

Fodor, Jerry A. 1983. *The modularity of mind*. Cambridge: MIT Press.

Hashimoto, Anne-Yue. 1971. *Mandarin syntactic structures*. Princeton: Princeton University.

Her, One-Soon. 1990. Grammatical functions and verb subcategorization in Mandarin Chinese. Ph.D. Thesis, Unversity of Hawaii.

Her, One-Soon. 2009. Unifying the long passive and the short passive: On the Bei construction in Taiwan Mandarin. *Language and Linguistics* 10 (3): 421–470.

Hsueh, F.S. 1989. The structural meaning of Ba and Bei constructions in Mandarin Chinese. In *Functionalism and Chinese grammar*, ed. James H-Y. Tai and F.S. Hsueh, 95–125. South Orange: Chinese Language Teachers Association.

Huang, Chu-Ren. 1992. Certainty in functional uncertainty. *Journal of Chinese Linguistics* 20 (2): 247–288.

Huang, Chu-Ren. 1993. Reverse long-distance dependency and functional uncertainty—The interpretation of Mandarin questions. In *Proceedings of the Korean Society for Language and Information Conference*, 111–120.

Huang, James C-T., Yen-Hui Audrey. Li, and Yafei Li. 2009. *The syntax of Chinese*. Cambridge: Cambridge University Press.

Jackendoff, Ray, and Jenny Audring. 2020. *The texture of the lexicon: Relational Morphology and the Parallel Architecture*. Oxford: Oxford University Press.

Kaplan, Ronald M, and Joan Bresnan. 1995. Lexical-functional grammar: A formal system for grammatical representation. In *Formal issues in lexical-functional grammar*, ed. Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell III, and Annie E Zaenen, 29–130. Stanford: CSLI Publications.

Kaplan, Ronald M., and John T. Maxwell III. 1996. *LFG grammar writer's workbench*. Technical report. Palo Alto: Xerox Palo Alto Research Center.

Kiss, Katalin É. 1995. Discourse configurational languages: Introduction. In *Discourse configurational languages*, ed. Katalin É Kiss, 3–27. Oxford: Oxford University Press.

Kit, Chunyu. 1998. Ba and Bei as multi-valence prepositions in Chinese—An LFG perspective. In *Studia Linguistica sinica*, ed. Benjamin K. Tsou, 497–522. City University of Hong Kong: Language Information Sciences Research Centre.

Lam, Olivia S.-C., One-Soon Her, Jing Chen, and Sophia Y.-M. Lee. to appear. LFG and Sinitic languages. In *The handbook of lexical functional grammar*, ed. Mary Dalrymple. Berlin: Language Science Press. https://langsci-press.org/catalog/book/312.

Li, Chao. 2019. Where to locate Mandarin localizers? *Journal of Chinese Linguistics* 47 (2): 345–370.

Li, Charles, and Sandra Thompson. 1976. Subject and topic: A new typology of language. In *Subject and topic*, ed. Charles Li, 457–489. New York: Academic Press.

Li, Fengqi. 2019. The mapping of thematic roles to grammatical functions in Mandarin BA construction denoting causation: From the perspective of Lexical Mapping Theory. *Journal of Language Teaching and Research* 10 (3): 618–626.

Li, Yen-Hui Audrey. 2006. Chinese Ba. In *The Wiley Blackwell companion to syntax*, ed. Martin Everaert and Henk Van Riemsdijk, 1:374–468. Oxford: Blackwell.

Lin, Dandan. 2019. 吴语平阳方言介词研究 [The research of prepositions in Pingyang dialect of Wu dialect area]. Master's Thesis, Hangzhou Normal University.

Lin, Jingxia, and John Newman. 2004. Wenzhou spoken corpus. http://www.artsrn.ualberta.ca/wenzhou/search.php. Accessed 13 Feb 2022.

Liu, Feng-Hsi. 1997. An aspectual analysis of Ba. *Journal of East Asian Linguistics* 6 (1): 51–99.

McCawley, James D. 1992. Justifying part-of-speech assignments in Mandarin Chinese. *Journal of Chinese linguistics* 20 (2): 211–246.

Müller, Stefan. 2020. *Grammatical theory: From transformational grammar to constraint-based approaches*. Berlin: Language Science Press.

Nie, Xiaowen, and Feng-Hsi. Liu. 2021. Disyllabic localizers in Modern Chinese. *Journal of Chinese Linguistics* 49 (1): 1–39. https://doi.org/10.1353/jcl.2021.0001.

Pan, Wuyun. 1997. 温州方言的动词谓语句 [Verbal predicates in Wenzhou dialect]. In 动词谓语句 *[Verbal predicates]*, ed. Rulong Li and Shuangqing Zhang, 58–75. Guangzhou: Jinan University Press.

Pan, Wuyun. 2002. 温州话的体和貌 [Aspect and mood in Wenzhounese]. In 著名中年语言学家自选集·潘悟云卷 *[Anthology of famous middle-aged linguists: Pan Wuyun]*, ed. Wuyun Pan, 181–209. Hefei: Anhui Education Publishing House.

Partee, Barbara H., Alice ter Meulen, and Robert E. Wall. 1990. *Mathematical methods in linguistics*. Dordrecht: Kluwer Academic Publishers.

Peyraube, Alain, and Thekla Wiebusch. 2021. New insights on the historical evolution of differential object marking (DOM) in Chinese. In *A typological approach to grammaticalization and lexicalization: East meets West*, ed. Janet Zhiqun Xing , 101–130. Berlin: De Gruyter Mouton.

Schachter, Paul. 1976. The subject in Philippine languages: Topic, actor, actor-topic, or none of the above? In *Subject and topic*, ed. Charles Li, 493–518. New York: Academic Press.

Scheer, Tobias. 2020. On the lexical character of intermodular communication. *Radical: A Journal of Phonology* 1: 183–239.

Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. *Phonology* 3: 371–405.

Selkirk, Elisabeth. 2011. The syntax-phonology interface. In *The handbook of phonological theory*, 2nd ed., ed. John Goldsmith, Jason Riggle, and Alan C. L.. Yu, 435–484. Oxford: Blackwell.

Sells, Peter. 1984. Syntax and semantics of resumptive pronouns. Ph.D. Thesis, University of Massachusetts Amherst.

Sells, Peter. 2013. Lexical-functional grammar. In *The Cambridge handbook of generative syntax*, ed. Marcel Den Dikken, 162–201. Cambridge: Cambridge University Press.

Shu, Chih-hsiang. 2018. Parameter preservation at the syntax-PF interface: The Ba construction revisited. *Studies in Chinese Linguistics* 39 (1): 29–66.

Spector, Ilona. 2009. Hebrew floating quantifiers. In *Proceedings of the LFG09 conference*, ed. Miriam Butt and Tracy Holloway King, 520–540. Stanford: CSLI Publications.

Sun, Chaofen. 2015. The grammaticalization of the BA construction: Cause and effect in a case of specialization. In *The Oxford handbook of Chinese Linguistics*, ed. William S-Y. Wang and Chaofen Sun, 429–442. Oxford: Oxford University Press.

Sun, Yangyu. 2018. A cartographic analysis of the syntactic structure of Mandarin Ba. *Studies in Chinese Linguistics* 39 (2): 127–154.

Sybesma, R. 1999. *The Mandarin VP*. Dordrecht: Springer.

Tan, Fu. 1991. Notion of subject in Chinese. Ph.D. Thesis, Stanford University.

Tsao, Feng-Fu. 1987. A topic-comment approach to the Ba construction. *Journal of Chinese Linguistics* 15 (1): 1–55.

Wang, Chi. 2017. 韵律对汉语致使结构谓语动词的制约 [Prosodic constraints on the predicate verbs of causative constructions in Mandarin Chinese]. *Studies in Prosodic Grammar* 2(1): 101–135.

Xu, Liejiong. 2015. Topic prominence. In *The Oxford handbook of Chinese Linguistics*, ed William S-Y. Wang and Chaofen Sun, 393–403. Oxford: Oxford University Press.

Xu, Lili. 2019. 浙江方言资源典藏：瑞安 *[A collection of Zhejiang dialects: Rui'an]*. Hangzhou: Zhejiang University Press.

Zhan, Weidong, Rui Guo, and Yirong Chen. 2003. The CCL corpus of Chinese texts: 700 million Chinese characters, the 11th Century B.C.–present. http://ccl.pku.edu.cn:8080/ccl_corpus. Accessed 17 Apr 2022.

Zhao, Chen. 2021. On middle applicatives: Evidence from Mandarin BA. *International Journal of Chinese Linguistics* 8 (1): 35–72.

Zhengzhang, Shangfang. 2008. 温州方言志 *[Wenzhou dialect]*. Beijing: Zhonghua Book Company.

Zhengzhang, Shangfang, and Wei Zheng. 2015. Wu dialect. In *The Oxford handbook of Chinese Linguistics*, ed. William S-Y. Wang and Chaofen Sun, 189–199. Oxford: Oxford University Press.