

Chinese Object Focus *VdeO* Cleft Sentences: A Corpus-Based Investigation

Alessia Iurato

Università Ca' Foscari Venezia, Italia

Laura Locatelli

Università degli Studi di Bergamo, Italia

Abstract This paper offers the first corpus-based study of the underexplored Chinese object-focus *VdeO* clefts (OFCs), which encode narrow focus on the object within a realized event. Using two large-scale corpora, we investigated their distribution with the 20 most frequent Chinese dynamic verbs, examining subject and object referents as well as focus patterns. Analyses show asymmetries in animacy, concreteness and referentiality, challenging earlier claims about object definiteness. OFCs predominantly convey non-contrastive focus, highlighting objects or supporting anaphoric recovery, while contrastive readings are infrequent and mostly implicit.

Keywords Chinese narrow-focus clefts. Object focus clefts. Focus realization. Semantic features. Corpus-based analysis.

Summary 1 Introduction. – 2 Chinese *VdeO* Cleft Sentences: Theoretical Background. – 3 The Object Focus Cleft. – 3.1 How the Adjacency Effect Relates to Focus in OFCs. – 3.2 Syntactic and Semantic Constraints on OFCs. – 3.3 Definiteness and Referential Properties of the Object. – 3.4 How *shì* and *de* Shape the OFC. – 3.5 Focus Realization and Discourse Functions of OFCs. – 4 The Present Study: Research Questions. – 4.1 Data and Methodology. – 4.2 Data Coding. – 5 Data Analysis and Interpretation. – 5.1 Frequency and Distribution of OFCs. – 5.2 Object and Subject Types. – 5.3 Correlation Between Verbs and Focus Types in OFCs. – 5.4 Focus Interpretation and Discourse Contexts of OFCs. – 6 General Discussion and Conclusions.



Peer review

Submitted 2025-12-19
Accepted 2026-06-19
Published 2026-07-30

Open access

© 2026 Iurato, Locatelli | © 4.0



DOI 10.30687/AnnOr/2385-3042/2026/62/004

1 Introduction

Chinese “proper clefts” (Paul, Whitman 2008, 424) with VdeO order (henceforth PCs) have drawn considerable attention in Chinese linguistics because of their distinctive structural properties.¹ These sentences are not fully parallel to clefts in other languages, such as English (Hole, Zimmermann 2013), and exhibit language-specific constraints (Hole 2011) that give rise to unique restrictions on their distribution and use.

PCs are used to establish narrow focus by highlighting a specific constituent that conveys new information within an event that is presupposed in the discourse (Iurato 2024a). In these syntactically marked constructions, focus is signaled by the morpheme *shì* 是, typically positioned adjacent to the focused element, and by *de* 的, which occurs postverbally immediately before the object (Hole 2011; Liu 2023).

PCs can be classified into three main types, depending on which constituent carries the focus: subject focus clefts, adjunct focus clefts, and object focus clefts (OFCs), in which the subject, a preverbal adjunct, or the postverbal object is highlighted, respectively, as illustrated in (1) (Paul, Whitman 2008; Liu 2023).² The most salient difference among the three types is that, in OFCs, the focused constituent occurs postverbally rather than in the canonical position adjacent to *shì*, a distribution shaped by specific syntactic constraints (Luo 2009).

- 1 a. 我们是骑自行车去的图书馆。
women shì qí zìxíngchē qù de túshūguǎn
 1PL SHI ride bike go DE library
 ‘It was *by bike* that we went to the library.’
 (Zhao 1979, 58)

Both Authors contributed equally to the present work. In accordance with academic conventions, the specific contributions of each Author are detailed below. A. Iurato is responsible for Sections 1, 2, 3, 3.1, 3.2, 3.3, 3.4, 3.5, 3.5.1, 3.5.2, 5.4. L. Locatelli is responsible for Sections 4, 4.1, 4.2, 5, 5.1, 5.2, 5.2.1, 5.2.2, 5.3, 6. Both Authors contributed to the writing, revision, and editing of the manuscript to an equal extent.

1 In this paper, the term ‘Chinese’ serves as a conventional abbreviation for ‘Standard Mandarin Chinese’. The glosses adhere to the guidelines of the Leipzig Glossing Rules, with the inclusion of *SP* = ‘structural particle’, *SHI* = ‘copular verb *shì* functioning as a focus marker’, *DE* = ‘sentence-final *de* in *shì...de* proper clefts’, *SFP* = ‘sentence final particle’.

2 Although the categories subject and object are theoretically problematic in Chinese (see Morbiato, Arcodia, Basciano 2020), this study adopts these labels following conventional descriptions of PCs. The term subject therefore refers to the preverbal, thematic constituent that functions as the participant initiating or carrying out the action, while object designates the postverbal element that undergoes or is affected by the event, whether direct, indirect, or locative in nature.

b. 是我去借的车。

shì wǒ qù jiē de chē
SHI 1SG go borrow DE car

‘It was *me* who went to borrow the car.’

(Zhao 1979, 59)

c. 张三是写的诗。

Zhāngsān shì xiě de shī
Zhangsan SHI write DE poem

‘It was *poems* that Zhangsan wrote.’

(Hole 2011, 1710)

Despite being a widely discussed topic in the literature, Chinese PCs remain notoriously challenging to analyze (Shyu 2015). Most studies conducted so far have adopted a generative approach (Cheng 2008; Paul, Whitman 2008; Hole 2011; Hole, Zimmermann 2013), while corpus-based research remains scarce. One notable exception is Xu (2014), who offered a general analysis of all studies involving the *shì...de* pattern, encompassing not only clefts but also copulative sentences with nominalization. Research from a pragmatic perspective is even more limited (e.g., Li 2008; Jing-Schmidt 2017).

What is clearly missing, then, is a comprehensive corpus-based perspective addressing focus realization within information structure and discourse contexts. Of the three main types of cleft constructions, existing literature focuses almost exclusively on adjunct-focus clefts, with only occasional attention to subject-focus clefts (e.g., Cheng 2008; Shyu 2015). OFCs, however, remain largely neglected: they are briefly mentioned in passing (e.g., Hole 2011; Long 2013), relegated to footnotes (e.g., Paul, Whitman 2008), and often absent from major reference grammars (Iurato 2024c). OFCs are the least explored cleft type both syntactically and in terms of information-structural interpretation, which contributes to difficulties in their recognition, comprehension, and use by both linguists and L2 learners of Chinese.

The present study addresses these gaps by offering the first large-scale corpus-based analysis of OFCs, examining their syntactic structure and the realization of focus within information structure and discourse contexts. As one of the few corpus-driven studies on Chinese clefts more broadly, it provides a systematic overview of this understudied construction. Based on large corpora of naturally occurring written data from L1 Chinese speakers, the study maps the distribution and usage patterns of OFCs, including verb-object co-occurrences, types of focus realized, and their potential correlations.

2 Chinese VdeO Cleft Sentences: Theoretical Background

Cleft constructions are widely regarded as a cross-linguistic strategy for signaling focus (Reeve 2013). They serve as tools for structuring information, typically exhibiting non-canonical word order and closely reflecting discourse-pragmatic organization (Hartmann, Veenstra 2013). In terms of information structure, clefts partition a sentence into a focused element, which introduces new information, and a presupposition encoding shared background knowledge (Halliday 1967; Lambrecht 1994; 2001). In English, this bipartition is realized in biclausal copular constructions, e.g., ‘It was in Paris that he fell in love’, where the cleft phrase (‘in Paris’) highlights new information against a presupposed event (‘he fell in love’) (Prince 1981). Across languages, clefts reorganize information structure in contrast to canonical orders, which rely on default syntax or prosody to encode focus (Lambrecht 2001; Hole, Zimmermann 2013).

In Chinese, this bipartite structure is implemented in VdeO clefts, usually organized as [NP1 *shì* Adv V-*de* NP2]. In this construction, *shì* serves both as a copula and as a focus marker, while the presupposed event is realized through a nominalized V-*de* clause (Hole, Zimmermann 2013). The constituent immediately following *shì* carries the cleft focus, highlighting new information, whereas the V-*de* headless relative clause represents presupposition, referring to information already established in the discourse (Long 2013).

By explicitly partitioning focus and presupposition, clefts represent a marked strategy for focus in Chinese, contrasting with canonical SVO sentences, where new information tends to appear clause-finally and receives default “end focus” (LaPolla 1990, 96). In unmarked SVO clauses, objects introducing new information generally occupy the sentence-final position, aligning naturally with this pattern. It is generally agreed that Chinese focus is associated with the postverbal domain: topical or presupposed elements occur preverbally, while focal elements appear postverbally (96-7). Thus, in canonical SVO sentences, focus gravitates to the clause-final slot and is highlighted prosodically, whereas clefts achieve prominence through structural partitioning. As a *wh*- and focus-in-situ language, Chinese also allows narrow focus to surface preverbally when a *wh*-element appears there, with presupposed material following the verb (Hole 2011; Paul 2015).

Within VdeO clefts, narrow focus can be syntactically realized only on NPs and PPs (Hole 2011). Focus is signaled by *shì*, which immediately precedes the focused element, while *de* appears postverbally before the object; their exact functions remain debated (see § 3.4). *Shì* can be omitted in some contexts, whereas *de* is obligatory (Cui, Sung 2021; Iurato 2024c). In terms of information structure, this construction allows the speaker to foreground a

constituent – in OFCs, the object of an event that has reached its endpoint in the past – against a presupposed background, making it a highly context-sensitive focus strategy (Iurato 2026).

3 The Object Focus Cleft

The following sections outline the key properties of OFCs. While some general features of PCs apply across all types (Paul, Whitman 2008; Hole 2011; Long 2013), the discussion here examines these phenomena from the perspective of OFCs, highlighting what is unique to this construction.

3.1 How the Adjacency Effect Relates to Focus in OFCs

Cross-linguistically, focus markers generally appear immediately before the constituent they highlight (Long 2013, 438). In Chinese, this generalization holds for subject and adjunct focus clefts, but it does not extend to OFCs. In the OFC pattern, the object – although it is the element that receives focus – remains in its postverbal position and is not fronted to a preverbal slot adjacent to *shì* (Hole 2011). The object's prominence is instead achieved prosodically rather than through linear adjacency to *shì* (Cheng 2008; Hole 2011).³

This differs from English *it*-clefts, where the phrase following the copula is normally the focus, typically a noun phrase (Reeve 2013). Chinese clefts – and OFCs in particular – exhibit a completely distinct structure. Long (2013) argues that PCs are biclausal copular constructions containing a headless relative clause nominalized by *de*. In such structures, the relative clause precedes the nominal head and occupies the position immediately adjacent to *shì*, while the focused object appears in its canonical postverbal slot. As a result, it is typically a subject or adjunct that appears next to the copula, providing the interpretive features needed to delimit the head noun. The adjacency effect in Chinese therefore targets a different syntactic domain from that of English clefts. Focus in OFCs is best conceived of as structurally, rather than linearly, encoded: the object may be embedded within the clause, while other constituents satisfy the adjacency condition. Long's account further underscores how cleft

³ One might tentatively suggest that the object is syntactically marked differently, being preceded by *-de* and no longer in the canonical verb-adjacent position. This is a plausible hypothesis, but it goes beyond the scope of the present study and is left for future investigation.

focus interacts with prosody and with the internal organization of relative clauses.

3.2 Syntactic and Semantic Constraints on OFCs

OFCs are subject to several types of restrictions. First, OFCs are subject to a series of TAM-related constraints. With respect to tense, they are incompatible with future-oriented temporal expressions or adverbials, since such elements would conflict with the presupposed nature of the event – an event understood as preceding the speech act (Iurato 2024c). Regarding aspect, the presuppositional component cannot host perfective markers (Liu 2023). Finally, as for modality, modal verbs are likewise excluded from the construction (Paul, Whitman 2008). Additionally, there are limitations on the semantic class of verbs: only dynamic verbs are allowed (Iurato 2024a).

Second, OFCs obey specific syntactic limitations. In Chinese, elements occurring in postverbal position, including objects, are resistant to clefting (Luo 2009). As a result, OFCs necessarily contain a postverbal object that itself cannot be clefted (Paul 2005; Hole 2011). According to Hole (2011, 1712), object focus arises through a postponed LF movement of the cleft-focus phrase after spell-out, which produces the characteristic configuration of OFCs.

3.3 Definiteness and Referential Properties of the Object

A key feature of OFCs is the requirement that the postverbal object be definite and specific in reference. Indefinite expressions such as the [one-CLF] pattern or combinations with the quantifier *jǐ* 几 (some) plus classifier are not acceptable in this context; the object requires clear definiteness marking, typically through a demonstrative (Cui, Sung 2021; Liu 2023). Long (2013) further argues that the object is not only definite but uniquely identifiable to both interlocutors, since it is anchored to a previously completed and discourse-accessible event. This anchoring to a completed, contextually recoverable situation ensures the object's strong referentiality, as it necessarily points back to a realized event preceding the speech time (Iurato 2024c). As a result, OFCs systematically encode events that are presupposed, completed, and temporally located prior to the utterance time.

3.4 How *shì* and *de* Shape the OFC

The functions of *shì* and *de* have been the subject of extensive debate. *Shì* is typically analyzed as a copula (Ross 1983; Cheng 2008; Paul, Whitman 2008), though some scholars treat it as an intensifying adverb (Han 2005) or a modal verb (Shi 1994). Following Hole (2011), Hole and Zimmermann (2013), Liu (2023), and Xu (2014), this study treats *shì* as a copula that also serves as a focus marker.

The role of *de* is even more contested. Since OFCs refer exclusively to past events, some studies (e.g., Song 1978; Simpson, Wu 2002) have proposed that *de* is an enclitic past-tense marker. However, this analysis is disputable, given that Chinese lacks a grammatical category of tense and instead relies on aspectual marking (Liu 2015). Others (e.g., Sugimura 1999; Liu 2023) interpret *de* as an aspectual marker, specifically as the anaphoric variant of the perfective marker *-le*. This study follows Iurato (2024c), who rejects this account on the grounds that the perfective interpretation of OFCs does not arise from *de* itself, but rather from the fact that the events described are already completed and presupposed, and therefore naturally precede the speech time.

The most widely endorsed view treats *de* as a nominalizer or relativizer of a headless relative clause (Hole 2011; Hole, Zimmermann 2013). In line with this perspective, we analyze OFCs as biclausal structures in which *de* nominalizes the presupposed event clause, while *shì* functions as a focus-marking copula. Although the object frequently appears post-*de* in these constructions, this ordering does not conflict with the nominalizing role of *de*: the post-*de* object contributes to the clause's referential content rather than serving as its internal head, reflecting the focus-driven reordering typical of cleft constructions (Long 2013). This analysis explains the perfective interpretation of OFCs by highlighting their characterization of completed and presupposed events; it also accounts for their past-time reading, since such events necessarily occur prior to the moment of utterance (Iurato 2026).

3.5 Focus Realization and Discourse Functions of OFCs

To date, no studies have provided a systematic or theoretically grounded account of the information-structural and discourse properties of OFCs. Since OFCs represent a subtype of PCs, the present discussion necessarily draws on the broader, albeit still limited, literature on PC focus interpretation. Existing accounts are heterogeneous and often only loosely connected to established focus-theoretical frameworks. Some studies (e.g., Shi 1994; Jing-Schmidt 2017) argue that PCs exclusively convey contrastive

focus, typically used to correct or clarify a prior assumption. Others (e.g., Li 2008) maintain that PCs may express either contrastive or non-contrastive focus. Xu (2014) proposes a more general emphatic function, suggesting that PCs serve as epistemic devices through which speakers signal stance or evaluation based on personal knowledge.

Following Iurato (2024c), this paper adopts the view that PCs, including OFCs, may express either contrastive or non-contrastive focus, consistent with the claim that clefts do not inherently encode contrastiveness (Rochemont 1986; Kiss 1998). The interpretation assigned to a cleft also depends on the notion of focus adopted, which varies considerably across theoretical frameworks. Given the lack of a coherent and empirically grounded account of focus realization in OFCs, the present study first establishes a working definition of focus for the purposes of this study and specifies the focus types guiding the subsequent analysis.

3.5.1 Defining Focus from an Information-Structure Perspective

Defining focus has long been a challenge, and no single account has gained universal acceptance. Within the information-structure literature (e.g., Jackendoff 1972; Rochemont 1986; Lambrecht 1994), focus is generally understood as the part of an utterance that conveys new or non-presupposed information. It corresponds to the element that the speaker chooses to highlight or foreground in order to direct the interlocutor's attention to it (Callies 2009). As such, the focal element represents the most salient piece of information in the communicative context – precisely the information through which the speaker pursues their communicative goals (Dik 1997).

Lambrecht's (1994, 213) pragmatic framework provides one of the most influential definitions of focus. According to him, focus is “the semantic component of a pragmatically structured position whereby the assertion differs from the presupposition” (Lambrecht 1994, 213). In Lambrecht's (1994, 52) words, the pragmatic assertion is “the proposition expressed by a sentence which the hearer is expected to know or take for granted as a result of hearing the sentence uttered”, whereas the pragmatic presupposition is “the set of propositions lexicogrammatically evoked in a sentence which the speaker assumes the hearer already knows or is ready to take for granted at the time the sentence is uttered”. Consequently, focus is defined as complementary to the presupposition, representing the portion of the utterance that conveys new or non-recoverable information to the listener. Crucially, focus cannot always be identified with a single constituent. As emphasized by Lambrecht (1994) and

Nikolaeva (2001), focality is not an inherent property of a referent but a relational category: it arises from the interaction between the utterance and its discourse context.

Focus can be further classified according to the scope of the focus domain. It is typically distinguished into three types: narrow focus, where the focus domain targets a single constituent (e.g., an object or an adjunct); predicate focus, where the focus domain encompasses the entire predicate; and sentence focus, where it extends to the whole clause (Lambrecht 2001). In the case of object focus clefts, given the intrinsic properties of PCs, the focus is necessarily narrow and targets only the object.

3.5.2 Contrastive and Non-Contrastive Focus

Focus has traditionally been considered contrastive (Garassino 2016, 192). However, numerous studies have shown that focus does not necessarily imply contrast, suggesting that contrast is not an inherent feature of focus (Halliday 1967; Kiss 1998; Rochemont 1986; Rooth 1992). Different taxonomies have been proposed to classify focus types (e.g., Vallduví, Vilkuna 1998). Among the most widely cited is the distinction between ‘information focus’ and ‘identification focus’ (Kiss 1998). For simplicity, this paper uses the general terms ‘contrastive’ focus and ‘non-contrastive focus’, depending on whether the focused element contrasts with other discourse-determined options. Contrastive focus involves a limited, identifiable set of alternatives, while non-contrastive focus involves an open set of potential referents (Callies 2009; Garassino 2016).

The concept of contrastiveness is central but debated. Halliday (1967, 206) defines contrastive as “contrary to some predicted or stated alternative”, while Chafe (1976) argues it requires: (i) background knowledge of the discourse, (ii) a limited set of plausible alternatives, and (iii) selection of one candidate to the exclusion of others. Repp (2010) similarly defines contrastiveness based on a closed, identifiable set of alternatives. However, as Kenesei (2006) points out, these conditions alone do not guarantee true contrastiveness. For a focus to be truly contrastive, it must satisfy the Exhaustiveness Condition – the focused constituent constitutes the complete set of entities determining the proposition’s truth – and the Exclusiveness Condition – only the focused entities determine the proposition’s truth, excluding all others (Chafe 1976; Declerck 1984; Kiss 1998). Sentences not meeting these conditions are non-contrastive.

Cleft sentences have traditionally been regarded as prototypically contrastive constructions (Declerck 1988; Lambrecht 2011). However, several studies have shown that they do not necessarily encode contrast (Prince 1978; Declerck 1984; König 1991; Kiss 1998;

Givón 2001). As noted by Miller (2006), cleft sentences may be either contrastive or non-contrastive, depending on the discourse context. This variability highlights the difficulty of determining when a focus should be analyzed as genuinely contrastive.

Contrastive focus is not merely a syntactic phenomenon: following Zimmermann (2008, 348), its interpretation depends on the speaker's assumptions about what the hearer considers likely or unlikely. Alternatives are therefore not only semantic (cf. Chafe 1976; Repp 2010) but also discourse-related, highlighting elements unexpected or in contrast with the hearer's presumed knowledge (see also Vallduví, Vilkuna 1998). Contrastive foci express a contrast between the information conveyed in asserting a constituent α and the hearer's assumed expectations, often using non-canonical grammatical forms to direct attention and update the common ground.

Contrastive focus can be explicit, occurring when syntactic or prosodic cues signal alternatives that are restricted and clearly defined within the communicative context (Biber et al. 1999; Garassino 2016), or implicit, when an identifiable set of alternatives is suggested but not overtly provided (Krifka 1999; Garassino 2016), or when the contrast arises purely from pragmatic inference and discourse context (Zimmermann 2008). Even when a focused element is syntactically identifiable, its contrastive interpretation may depend on background information, the alternatives the speaker assumes the hearer expects, or the speaker's intended emphasis (Zimmermann 2008). This aligns with Lambrecht's (1994, 58) principle that "focality is not a property of a referent as such but rather a relational category", emerging from the interaction between utterance and discourse context.

The interaction between contrastive focus and wh-questions illustrates these principles. In wh-questions and in their ordinary, informative answers, contrastive marking is typically absent, since the requested information is simply supplied and no contrast is intended. Contrastive focus emerges only in marked responses, such as corrections or unexpected answers, where the speaker explicitly challenges or revises the hearer's assumptions (Zimmermann 2008). In such cases, even answers to wh-questions can display contrastive marking, for instance to reject likely expectations, to signal that only a single individual satisfies the predicate, or to contrast the actual referent with an alternative previously assumed by the hearer (Zimmermann 2008).

In this study, we follow Zimmermann (2008), according to which contrastive focus targets information that is unexpected or that stands in contrast to the hearer's presumed knowledge, whether this contrast is made explicit in the discourse or only implicitly inferred.

Whenever the focused element does not activate such a contrastive interpretation, it is classified as non-contrastive focus.⁴

To account for non-contrastive focus, we adopt Korzen's (2014, 232) proposal, according to which it fulfils the basic function of marking the "original" focus. Its role is to direct the listener's attention to a specific piece of information concerning an event, which is completed in the OFC, without introducing any contrast.

Drawing on the literature (Berretta 1994; Zimmermann 2008), completive, anaphoric, and information-seeking functions are identified as key manifestations of non-contrastive focus in cleft constructions: (i) completive, to highlight a relevant aspect in order to complete or refine the set of information available to the interlocutor; (ii) anaphoric, to reintroduce previously mentioned information and restore it to discourse prominence; and (iii) information-seeking (typically in *wh*-questions), to request specific information about an event or element whose properties are not fully known.

In the present study (see § 5.4), these distinctions are not treated as independent theoretical categories, but are operationalized in the corpus analysis as three recurrent discourse contexts of OFCs: narrative-descriptive contexts (broadly corresponding to completive uses), anaphoric contexts, and *wh*-question contexts (corresponding to information-seeking uses).

The classification adopted in this study therefore integrates both focus interpretation and discourse context, allowing us to determine whether a given OFC instantiates contrastive or non-contrastive focus on the basis of its full contextual environment.

4 The Present Study: Research Questions

Given the aims and rationale outlined in Section 1, and in light of the existing literature and remaining gaps, the present study is guided by the following research questions:

- RQ1: How are OFCs distributed across the 20 most frequent Chinese dynamic verbs, and which text type emerges as the dominant for their occurrence?

⁴ A full understanding of communicative purposes and focus often requires attention to the prosodic level. Many studies of clefts have examined prosody, as pitch accent and intonation provide key cues, signaling a contextually salient set of alternatives (Bolinger 1961; Rooth 1992) and helping distinguish contrastive from non-contrastive focus, especially when syntactic or morphological markers are absent. However, this aspect lies beyond the scope of the present study, which is based on written data and does not analyze prosodic features.

- RQ2: What semantic and referential properties characterize the object and subject found in OFCs, and how do these features correlate with the verbs that occur in the construction?
- RQ3: Are OFCs predominantly used to convey contrastive or non-contrastive focus, and is there a correlation between the type of focus realized and the verbs used?

4.1 Data and Methodology

Before examining which dynamic verbs can occur in OFCs, we identified the 20 most frequently used dynamic verbs in Chinese.⁵ Frequency information was obtained from two sources: the Wiktionary frequency lists (1-1,000 words),⁶ based on consolidated corpora (Bajčetić, Declerck 2022), and the Chinese Toolbox, which generates frequency data from the Sinica Balanced Corpus of Modern Chinese.⁷ Combining these sources allowed us to identify a representative set of high-frequency dynamic verbs that reflect common usage in naturally occurring Chinese texts.

The selected verbs were then classified according to the semantic verb classes identified by Levin (1993). In a few cases, the classification was slightly adjusted to better reflect the verbs' meaning and usage in contemporary Chinese. Specifically, verbs of creation such as *zuò* 做 (make/do) were retained in their original category, whereas verbs like *xiě* 写 (write) and *chàng* 唱 (sing), originally listed under message transfer or manner-of-speaking verbs in Levin's framework, were reclassified as creation verbs, as they involve producing a concrete output (Chang, Chen, Huang 2000). Likewise, *zuò* 坐 (take a vehicle), which Levin (1993) categorizes as a verb of motion using a vehicle, was reassigned to instrumental locomotion verbs, following Chang, Chen and Huang (2000), since the agent undergoes motion while the object denotes the instrument enabling the movement. Table 1 presents the 20 selected verbs along with their adapted taxonomy used in this study.

5 The decision to limit the analysis to a controlled set of 20 high-frequency dynamic verbs was motivated by methodological considerations. A broad corpus query using a general pattern such as '是 V 的 O' would have generated a large number of heterogeneous and often irrelevant results, including constructions that fall outside the operational definition of OFCs adopted in this study. To avoid this noise and ensure analytical consistency, we therefore opted for a restricted, frequency-based selection of verbs rather than a fully corpus-driven extraction.

6 For the Wiktionary frequency lists, see: <https://www.wiktionary.org/>.

7 For the Chinese Toolbox, which provides frequency data from the Sinica Balanced Corpus of Modern Chinese, see: <https://asbc.iis.sinica.edu.tw/>.

Table 1 Dynamic verbs selected for OFC analysis

Verb	Semantic verb class
1. <i>chī</i> 吃 (eat)	Verbs of ingesting
2. <i>hē</i> 喝 (drink)	
3. <i>zuò</i> 做 (make/do)	Verbs of creation
4. <i>xiě</i> 写 (write)	
5. <i>chàng</i> 唱 (sing)	Verbs of transfer of message
6. <i>wèn</i> 问 (ask)	
7. <i>mǎi</i> 买 (buy)	Verbs of change of possession
8. <i>mài</i> 卖 (sell)	
9. <i>gěi</i> 给 (give)	Learn verbs
10. <i>xué</i> 学 (study)	
11. <i>kàn</i> 看 (read)	Verbs of perception
12. <i>tīng</i> 听 (listen)	
13. <i>ná</i> 拿 (take/hold)	Verbs of carrying
14. <i>dài</i> 带 (carry/bring/take along)	
15. <i>lái</i> 来 (come)	Verbs of motion
16. <i>qù</i> 去 (go)	
17. <i>zǒu</i> 走 (walk)	
18. <i>jìn</i> 进 (enter)	
19. <i>zuò</i> 坐 (take a vehicle)	Instrumental locomotion verbs
20. <i>zhǎo</i> 找 (search/look for)	Verbs of searching

After selecting the verbs, we isolated the sample using two large-scale, annotated, genre-balanced Chinese corpora, comparable in size and diversity, supporting robust cross-corpus comparison. On the one hand, the BCC corpus offers multi-genre materials, including academic texts, newspapers, literature, and social media,⁸ while the zhTenTen Simplified Chinese corpus, available on Sketch Engine, provides web-based language from forums, commercial sites, and social media.⁹

For each of the 20 verbs listed above, we extracted a sample of 600 sentences from each corpus, using queries targeting the verb within a OFC, such as *shì chī de* 是吃的. We intentionally excluded more complex queries with part-of-speech tags in which the construction is immediately followed by a noun, as is commonly the case in OFCs, since this would have limited our ability to observe a wider range of structural patterns. When fewer than 600 sentences were available for a given verb, the full set of available sentences was included.

⁸ The BCC corpus is accessible at: <https://bcc.blcu.edu.cn/>.

⁹ The zhTenTen Simplified Chinese corpus is accessible at: <https://www.sketchengine.eu/zhtenten-chinese-corpus/>.

The final dataset therefore comprises 13,176 sentences in total, also including their immediate context, with 2,209 from the BCC and 10,967 from the zhTenTen.

The next step consisted of an in-depth data cleaning process. During this stage, we independently isolated only those instances that qualified as OFCs, reaching an inter-rater agreement of $\kappa = .93$. The resulting cleaned dataset, which forms the basis of the analyses presented in this study, includes 4,585 sentences, with 1,653 from the BCC and 2,932 from the zhTenTen.

4.2 Data Coding

For the analyses, the final dataset was fully annotated along four main dimensions: object, subject, focus, and text type. Regarding the first two, NPs functioning as objects and subjects in each OFC were annotated for animacy and concreteness, two semantic properties that capture, respectively, the extent to which a referent is perceived as alive and sentient (Comrie 1989), and the degree to which it can be experienced through the five senses (Brysbaert, Warriner, Kuperman 2014). Following Locatelli (2026), both properties were treated as continuous, rather than categorical. In particular, animacy was calculated as the mean of four components: ‘being alive’, ‘looking like a human’, ‘having mobility’, and ‘having agency’, each coded on a binary scale, with 1 indicating the absence and 2 the presence of the feature. For example, *shípǐn* 食品 (food) scores 1 on all four factors, yielding an average of 1, which corresponds to complete inanimacy. By contrast, *rénmen* 人们 (people) scores 2 across all dimensions, reflecting not only full animacy, but also humanness. Similarly, concreteness values were derived from the mean ratings in Brysbaert, Warriner and Kuperman (2014), which range from 1 (highly abstract) to 5 (highly concrete). Each subject and object referent in our sample was assigned the rating of its English lexical counterpart.¹⁰ For instance, the object NP *niúǎi* 牛奶, corresponding to the English term ‘milk’ in the dataset, received a concreteness rating of 4.9, indicating a highly concrete entity.

In addition, since previous research has shown that referentiality plays a crucial role in Chinese clefts (see Hole 2011; Iurato 2024a), each object and subject was also annotated for this property.

10 As discussed by Locatelli (2026), while translation-based annotation is not without its limitations, it was adopted here because no comprehensive Chinese datasets are available that are comparable to the English norms by Brysbaert, Warriner, and Kuperman (2014). Accordingly, each NP was translated for annotation with attention to maintaining its contextual meaning, thereby minimizing potential ambiguities deriving from polysemy.

Following Chen (2009), linguistic referentiality is understood as not only a semantic notion but fundamentally a pragmatic one: beyond signaling the existence of an entity in its decontextualized meaning, referential expressions are primarily used speaker-dependently to track or identify an entity within the universe of discourse. Therefore, annotation proceeded in two stages. We first assigned referential values on the basis of grammatical form, relying on categories commonly treated as referential, and consequently definite or indefinite, or non-referential in the literature (see Chen 2004; Jiang 2015). Proper names, personal pronouns, and NPs introduced by demonstratives were classified as definite (*def*), while numeral constructions – including those with *yī* 一 (one), *xiē* 些 (some) – as well as bare NPs were provisionally classified as indefinite (*indef*). In a second step, we examined each OFC in its immediate discourse context to identify cases in which expressions that appeared indefinite based on form alone in fact referred to a specific, discourse-presupposed entity. These instances were therefore reclassified accordingly.

As for the focus dimension, each sentence in the corpus was analyzed individually within its full discourse context to determine whether it conveyed contrastive or non-contrastive focus, according to the notion of contrastivity outlined in Section 3.5.2. In addition, for each instance, we coded whether the focus was explicit, signaled by syntactic or morphological markers, or implicit, inferred from the pragmatic and contextual cues in the discourse.

Finally, each concordance line was annotated for text type, understood as the genre in which the OFC occurs. The annotation followed a combined top-down and bottom-up procedure. When available, we first relied on the genre labels supplied in the corpora metadata – academic (A; *xuéshù* 学术), literary (L; *wénxué* 文学), newspaper (N; *xīnwén* 新闻), and social media (S; *shèjiāo méitǐ* 社交媒体). These initial categories were then refined and expanded to include administrative-institutional (AD; *xíngzhèng guāngfāng* 行政官方) and informational-technology (IT; *xìnxī jìshù* 信息技术) texts.

5 Data Analysis and Interpretation

This section presents the results of the analysis of OFCs. It begins by describing their overall frequency and distribution. The analysis then examines the distribution of object and subject types, followed by the relationship between verbal categories and focus types. Finally, it considers the discourse contexts in which OFCs occur and their contribution to focus interpretation.

5.1 Frequency and Distribution of OFCs

The distribution of OFCs exhibits clear verb-level asymmetries. Across both corpora, *zǒu* emerges as the most frequent verb, followed closely by *zuò* in the zhTenTen and *mǎi* in the BCC. In the BCC, besides *zuò*¹¹, which ranks fifth overall, other high-frequency verbs include *kàn*, *qù*, *chàng*, and *xiě*. A similar pattern is observed in the zhTenTen, with all these verbs occupying top positions in the frequency ranking. Notably, *lái* is entirely absent from OFCs in both corpora, distinguishing it sharply from its motion counterpart *qù*, which appears consistently in both datasets. Other verbs are also generally infrequent in OFCs across both corpora, including *tīng* and *wèn*. Conversely, several display more divergent patterns, like *jìn* and *dài*, which appear rarely in the zhTenTen, whereas *xué* and *zhǎo* are scarce or absent in the BCC. In particular, *zhǎo* is unattested in the BCC but ranks as the fourth most frequent verb in the zhTenTen, suggesting corpus-specific differences in the deployment of certain verbs. However, a Spearman correlation analysis of verb-wise OFC frequencies revealed a significant positive association between the two corpora ($\rho = .647$, $p = .002$), indicating that the relative ranking of verbs is largely consistent despite overall frequency differences. This suggests a stable distributional pattern of OFCs, with high-frequency items primarily associated with verbs of motion and creation, while verbs denoting perception or change of possession display greater variability across corpora. Table 2 presents the normalized frequencies of OFCs by verb.¹²

Table 2 Frequency of OFCs by verb

Verb	BCC	zhTenTen	Verb	BCC	zhTenTen
1. <i>chàng</i> 唱	0.776	2.3464	11. <i>ná</i> 拿	0.339	0.8408
2. <i>chī</i> 吃	0.533	2.0865	12. <i>qù</i> 去	1.163	3.0377
3. <i>dài</i> 带	0.194	0.5941	13. <i>wèn</i> 问	0.097	0.8609
4. <i>gěi</i> 给	0.097	0.8592	14. <i>tīng</i> 听	0.145	0.6337
5. <i>hē</i> 喝	0.582	0.9602	15. <i>xué</i> 学	0.097	2.2031
6. <i>jìn</i> 进	0.145	0.2844	16. <i>xiě</i> 写	0.679	2.3315

11 Since the characters 做 and 坐 share the same pinyin and tone, they are distinguished in the text as *zuò1* and *zuò2*, respectively. In the figures, due to technical limitations in R Studio (ver. 2025.09.1+401) that prevent tone-marking, the two verbs appear simply as *zuo1* and *zuo2*. The same applies to *mǎi* 买 and *mài* 卖 which, despite differing in tone, are shown in the graphs in atonal pinyin as *mai1* and *mai2*.

12 The frequencies reported represent estimated occurrences, calculated based on the cleaned sample of OFCs and adjusted proportionally to reflect the total number of concordances in the corpus. To enable comparison across corpora of different sizes, all frequency values were normalized to a standard base of 1,000,000 tokens.

Verb	BCC	zhTenTen	Verb	BCC	zhTenTen
7. kàn 看	1.260	2.5305	17. zuò 做	1.018	6.1719
8. lái 来	0.000	0.0000	18. zuò 坐	0.436	2.5165
9. mǎi 买	1.309	5.9415	19. zǒu 走	1.939	6.8845
10. mài 卖	0.145	1.0216	20. zhǎo 找	0.000	3.2909

Despite the positive correlation between verb rankings in the two corpora, OFCs' normalized frequencies differ substantially. A paired t-test on verb-wise frequencies confirms this tendency, showing that OFCs occur significantly more often in the zhTenTen than in the BCC ($t(19) = 4.83$, $p < .001$), with a mean difference of 1.72 per million (95% CI: 0.98-2.47). Verb-specific z-scores reinforce this pattern: *zuò1* ($z = 1.92$), *zǒu* ($z = 1.66$), *zhǎo* ($z = 1.81$), and *mǎi* ($z = 1.72$) are strongly overrepresented in the zhTenTen, whereas only a few verbs, including *hē*, *jìn*, and *dài*, show minimal divergence ($z < 1$). These findings indicate that, although the relative rank of verbs is largely preserved, the larger size and greater diversity of the zhTenTen contribute to systematically higher frequencies. Unlike the BCC, which is smaller and less varied, the zhTenTen encompasses a broad range of web-based texts – forums, commercial sites, social media, and IT/informational registers – enhancing the likelihood of capturing OFCs. This pattern is also evident in Figure 1, which clearly shows that the distribution of OFCs by verb across different text types is markedly more heterogeneous in the zhTenTen compared to the BCC.

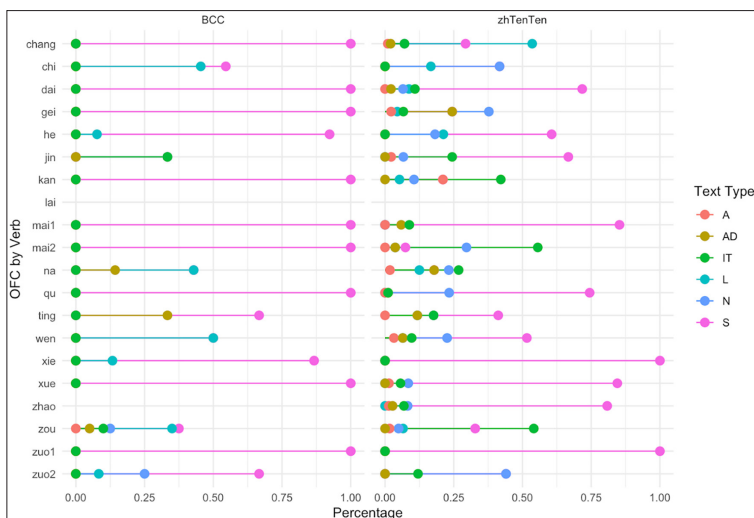


Figure 1 Verb-wise distribution of OFCs across text types. Chart created by the Authors

Although OFCs show greater overall dispersion in the zhTenTen, text-type analysis points to a clear similarity between the two corpora. In both datasets, OFCs are primarily attested in social media (S), reflecting their preference for more colloquial and informal registers. This pattern is confirmed by a chi-square test on the overall distribution of text types, which revealed no significant differences between the corpora ($X^2 = 5.17$, $df = 5$, $p = .396$), and further reinforced by a chi-square comparison contrasting social media with all other text types (S vs. NonS), which also showed no significant difference ($X^2 = 1.59$, $df = 1$, $p = .21$). These results indicate that, despite differences in absolute frequencies and the broader coverage of the zhTenTen, OFCs for most verbs are predominantly found in genres resembling spoken, informal communication, thereby underscoring a core characteristic of this structure and supporting previous research on Chinese PCs (see Iurato 2024b; Mai, Yuan 2016).

5.2 Object and Subject Types

To assess the semantic and referential properties of both subjects and objects in terms of animacy, concreteness, and referentiality, descriptive statistics – including means, standard deviations, and ranges – were computed for each parameter. While verb-specific variability is analyzed in the following subsections, this section focuses on potential differences between the two arguments across the three annotated parameters.

Results show a clear asymmetry: subjects are overwhelmingly animate, mostly human (mean animacy = 1.89, $SD = 0.21$), moderately concrete (mean = 3.35, $SD = 0.28$), and highly referential (mean = 1.00, $SD = 0.02$), indicating highly identifiable discourse entities. Objects are largely inanimate (mean animacy = 1.06, $SD = 0.11$), slightly more concrete on average (mean = 3.96, $SD = 0.58$), and exhibit greater variability in concreteness and referentiality (mean reference = 0.83, $SD = 0.28$). All differences are statistically significant, confirming systematic semantic and referential distinctions between NP1 and NP2 in OFCs.

Importantly, these patterns are consistent across corpora. In both the zhTenTen and BCC, subjects remain highly animate, moderately concrete, and strongly referential, whereas objects are inanimate, concrete, and moderately referential. Wilcoxon signed-rank tests confirm that differences between NP1 and NP2 are statistically significant for all three parameters in both corpora.

A preliminary verb-level analysis further clarifies these patterns. Some verbs, including *hē*, *mǎi*, *mài*, *qù*, *tīng*, and *zuò2*, show nearly uniform NP1 animacy, whereas others, like *jìn*, *xiě*, and *zǒu*, display higher variability (SDs 0.36-0.57, ranges up to 0.8). Object properties

are generally more variable, especially for referentiality and concreteness, as seen with verbs like *gěi*, *tīng*, and *xué*, whereas verbs such as *chàng* and *zǒu* select NP2s that are more consistently concrete and low in referentiality. Figure 2 illustrates these subject-object differences across all three parameters for each verb.

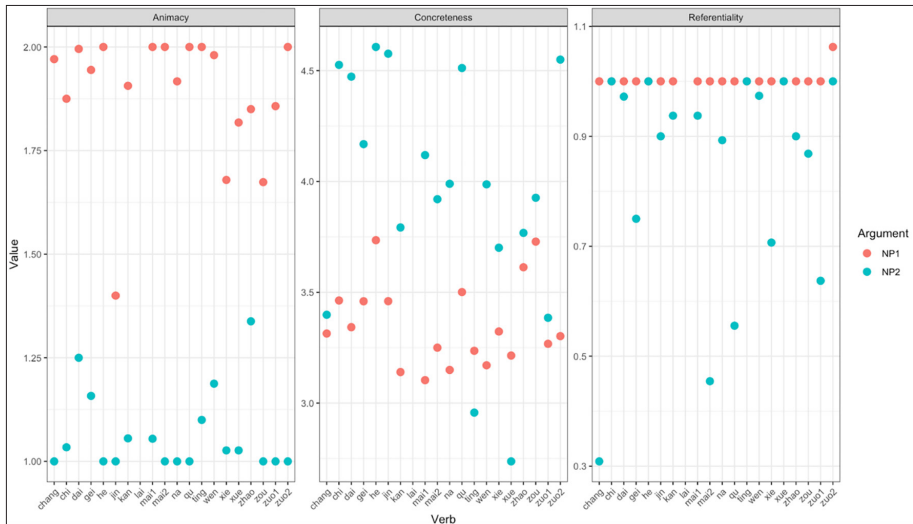


Figure 2 Animacy, concreteness, and referentiality of subjects and objects in OFCs per verb. Chart created by the Authors. As described in Section 4.2, the values for animacy and concreteness represent mean scores, ranging from 1 to 2, and 1 to 5, respectively. For referentiality, the figure depicts the proportion of referential entities for each verb, scaled from 0 to 1

5.2.1 Object Types in OFCs

We begin the verb-level analysis from the object, the core argument in an OFC. A series of ANOVAs reveals that the nature of the object differs significantly across verbs for two of the three annotated parameters. Concreteness ($F = 22.64$, $p < 2e-16$) and referentiality ($F = 9.718$, $p < 2e-16$) show strong verb-dependent variation, whereas animacy does not ($F = 1.204$, $p = .273$). This indicates that, although verbs differ markedly in the kinds of entities they select with respect to abstractness and referential status, they tend to select objects with comparable levels of animacy.

As the only parameter that does not vary significantly across verbs, animacy shows a remarkable degree of stability within OFCs, with values consistently cluster around 1, corresponding to fully inanimate entities. Most verbs exhibit no internal variability ($SD = 0$), underscoring the uniformity of this dimension. Examples

(2) and (3) clearly illustrate this pattern.¹³ In (2), the verb *hē* selects the inanimate object *zhè-kuǎn nǎifěn* 这款奶粉 (this formula/milk powder), while in (3), the verb *zuò* similarly takes the inanimate NP2 *huǒchē* 火车 (train), aligning perfectly with the general tendency of OFCs to feature entities that completely lack animacy as their objects.

- 2 推荐英国牛栏奶粉, 我家宝宝段奶后一直都是喝的这款奶粉, 在百儿母婴商城网站代购的。(zhTenTen)

<i>wǒ</i>	<i>jiā</i>	<i>bǎobao</i>	<i>duànnǎi</i>	<i>hòu</i>	<i>yīzhí</i>	<i>dōu</i>
1SG	family	baby	wean	after	always	all
<i>shì</i>	<i>hē</i>	<i>de</i>	<i>zhè-kuǎn</i>	<i>nǎifěn</i>		
SHI	<i>hē</i>	DE	this-CLF	formula		

‘I recommend Cow & Gate milk powder from the UK. Ever since my baby was weaned, it has been this formula that he has always drunk, which we bought through the Bai’er Mother & Baby online store.’

- 3 这辈子还没坐过飞机, 就连玩意大利也是坐的火车。(BCC)

<i>jiù</i>	<i>lián</i>	<i>wán</i>	<i>Yìdàlì</i>	<i>yě</i>	<i>shì</i>	<i>zuò</i>	<i>de</i>	<i>huǒchē</i>
ADV	even	play	Italy	also	SHI	take	DE	train

‘I’ve never flown in my life; even when I went to Italy, it was the train that I took.’

However, a small subset of verbs allows for animate NP2s. One notable example is the verb *dài*, which can take humans as objects when the action involves physically accompanying or guiding someone. In (4) the speaker uses *dài* to indicate that they brought two professors along on a tour, selecting a fully animate and human object, i.e., *liǎng-ge jiàoshòu* 两个教授.

- 4 导游李萍是个非常不错的小姑娘, 讲解很熟练, 历史知识也非常丰富, 我是带的两个教授去的, 教授对小李的讲解都竖大拇指, 非常好。(zhTenTen)

<i>wǒ</i>	<i>shì</i>	<i>dài</i>	<i>de</i>	<i>liǎng-ge</i>	<i>jiàoshòu</i>	<i>qù</i>	<i>de</i>
1SG	SHI	bring.along	DE	two-CLF	professor	go	SFP

‘The tour guide Li Ping is a really excellent young lady. Her explanations are very skilled, and her knowledge of history is extremely rich. It was two professors that I brought along, and they both gave Li a thumbs-up for her explanations – absolutely fantastic.’

¹³ From this point onward, examples present the OFCs together with their immediate surrounding context. For the sake of clarity and economy of presentation, interlinear glosses are provided only for the OFC itself.

A similar pattern is observed with dative verbs such as *wèn* and *gěi*, which, in addition to taking a direct object, typically include an indirect object representing the recipient of the action (Chin 2010). Unlike the direct NP, which commonly denotes the thing being acted upon, the indirect object in ditransitive constructions is typically animate, as only sentient entities are able to receive. In (5) illustrates this, with *shūshu* 叔叔 (uncle) satisfying the semantic requirement of the verb for a sentient recipient. Interestingly, *Siri* – though not biologically animate – is also treated here as capable of receiving the action, acquiring animacy for the speaker.

5 你是问的叔叔还是Siri啊? (BCC)

nǐ shì wèn de shūshu hái shì Siri a
2SG SHI ask DE uncle or Siri SFP
'Is it your uncle or Siri that you asked?'

In contrast to animacy, verbs exhibit considerable variability in the concreteness of their objects. Some verbs show strong internal homogeneity, such as *chàng* (SD = 0), *xiě* (SD = 0.174), and *xué* (SD = 0), whereas others permit a much broader range of object types, including *chī* (SD = 1.89), *mài* (SD = 2.12), and *zuò* (SD = 1.47). Across all verbs, mean concreteness values span from 2.6 to 4.8, indicating that objects in OFCs vary from relatively abstract to highly concrete. The verb *xué* consistently selects the least concrete objects (mean = 2.6 in the BCC; 2.8 in zhTenTen), immediately followed by *tīng* (mean = 2.6 in the BCC; 3.2 in zhTenTen). Examples are provided in (6) and (7).

6 省长很宽容地笑了笑, “我也是学的物理专业, 但以后这三十年, 我都差不多忘光了”。(BCC)

wǒ yě shì xué de wùlǐ zhuānyè
1SG also SHI study DE physics major
'The governor smiled tolerantly. "It was physics that I also majored in, but over the past thirty years, I've pretty much forgotten it all".'

7 刚好教练以前是4S店上班的, 所以当时每次选的几款车也是听他的建议。(zhTenTen)

yě shì tīng de tā de jiànyì
also SHI listen DE 3SG SP advice
'Coincidentally, my instructor used to work at a 4S dealership so, when selecting a few models, it was his advice that I followed.'

While verbs such as *zuò1*, *xiě*, *kàn*, and *chàng* tend to select NP2s of moderate concreteness in OFCs, a small subset of predicates stands out for their strong preference for highly concrete objects. Among these, the motion verbs *qù* and *jìn* (mean = 4.4 and 4.8 in the BCC; 4.5 and 4.3 in zhTenTen, respectively), as well as the instrumental locomotion verb *zuò2* (mean = 4.6 in the BCC; 4.5 in zhTenTen), typically combine with locative objects denoting physically bounded places or tangible means of transport. As such, these NP2s consistently show a high degree of concreteness, as in (8).

- 8 汪永富第一次找尤金是去的办公室, 这一次求人心切, 等不及尤金上班, 直接闯到他的家里去。(BCC)

Wāng.Yǒngfù dìyī-cì zhǎo Yóujīn shì qù de bàngōngshì
Wang.Yongfu first-time look.for Youjin SHI go DE office

‘The first time Wang Yongfu went to look for You Jin, it was in his office that he went. But this time, anxious to ask for help and unable to wait for You Jin to go to work, he went straight to his home.’

By contrast, the other motion verb *zǒu* exhibits a more heterogeneous profile. Although it does select relatively concrete objects – such as *qiánmén* 前门 (the front gate) (BCC) or *zhè-tiáo lù* 这条路 (this road) (zhTenTen) – it also occurs frequently in figurative uses where the NP2 denotes not a physical location but a metaphorical path or conceptual orientation. Representative examples include *shèhuìzhǔyì dàolù* 社会主义道路 (the socialist path) (zhTenTen) and ROCK *fēnggé* 风格 (a rock style) (BCC).

The final parameter, referentiality, demonstrates a pronounced preference for definite and discourse-identifiable NP2s, with an internal proportion often exceeding 80% of total occurrences for a given verb. While personal pronouns are nearly absent in the dataset, some NP2s appear as proper names, which unambiguously signal an identifiable referent shared by both interlocutors, as illustrated in (9).

- 9 认识钟医生11年了, 生大宝的时候产检, 生产都是看的钟医生。(zhTenTen)
chǎnjiǎn shēngchǎn dōu shì kàn de Zhōng Yīshēng
check.up delivery all SHI look.for DE Zhong doctor
‘I have known Dr. Zhong for eleven years. When I had my first child, it was Dr. Zhong that I saw for all my prenatal check-ups and the delivery.’

A frequent case of explicit semantic definiteness is also represented by NP2s introduced by demonstratives. For verbs such as *chī*, *hē*, *mǎi*, *wèn*, *zuò2*, and *jìn*, demonstrative-marked NP2s account for between 25% and 70% of all occurrences. As Chen (2004) argues, a demonstrative functions as a determiner of definiteness, often serving

a deictic role by locating and identifying entities with respect to the spatio-temporal context of the discourse. Example (10) illustrates this pattern, where the object *zhè-ge chá* 这个茶 (this tea) provides a specific, identifiable referent for the action of drinking.

- 10 《铁齿铜牙纪晓岚4》乾隆到柳树沟找梅香姑娘就是喝的这个茶。(zhTenTen)

jiù shì hē de zhè-ge chá
ADV SHI drink DE this-CLF tea

‘In “Iron Teeth Bronze Teeth Ji Xiaolan 4”, when Emperor Qianlong goes to Liushu Gou to look for the girl Meixiang, it is this tea that he drinks.’

Most objects, however, do not fall into explicit categories that unambiguously signal definiteness, as discussed in Section 4.2. Nevertheless, 43% of NP2s are still annotated as definite because they consist of nouns modified by nominal determiners that uniquely or specifically identify the referent in context. These modifiers function similarly to demonstratives by restricting the reference of the noun to a single entity or type within the discourse. In (11), the object *pèngpèngchē zhuānyòng jiàzhào* 碰碰车专用驾照 (bumper car license) is interpreted as definite, because the modifiers preceding *jiàzhào* 驾照 (car license) restrict its reference to a single, contextually identifiable type of license, known to both the interlocutors.

- 11 车都冒烟了还不停, 难道这哥们是拿的碰碰车专用驾照?(BCC)

nándào zhè gēmen shì ná de pèngpèngchē zhuānyòng jiàzhào
could.it.be this guy SHI bring DE bumper.car special license

‘The car was smoking and still didn’t stop, could it be that this guy actually got a bumper car license?’

In instances where the object is realized as a bare NP, contextual information was essential for annotating it as definite, indefinite, or non-referential. In research on Chinese language, it is well-known that bare NPs are typically non-identifiable when appearing postverbally (see Chen 2004). However, Cummins (1998, 578) argues that bare nouns in sentence-final positions are generally interpreted as definite only when the entity is “at hand in context or in memory”. In the vast majority of cases examined in this study, bare NP objects were annotated as definite, because the discourse context rendered the referent uniquely identifiable. For example, the OFC in (12) introduces the object *mǐxiàn* 米线 (rice noodles), establishing a clear discourse anchor, in which (i) the noodles in question are those served at that specific restaurant on that morning, (ii) they are already familiar to the speaker or listener within the ongoing narrative, and (iii) no other competing referents exist within the local discourse context. In this

case, the bare NP object achieves definiteness through pragmatic and contextual cues.

- 12 点了在旅游书上面介绍的几个凤凰的特色菜，还比较喜欢那个“血粑烧鸭”，他们那的糍粑都是黑黑的，好像是用鸭血裹了的！早上在那里当然是吃的米线，感觉没有以前学校旁边的“湖南米线”够味，不过他们的“马子”好多。
(zhTenTen)

zǎoshang zài nàlǐ dāngrán shì chī de mǐxiàn
morning at there of.course SHI eat DE rice.noodle

‘I ordered several of Phoenix’s specialty dishes featured in the travel guide, and I especially liked the ‘blood rice cake roasted duck’. Their rice cakes are all dark in color, as if coated in duck blood! In the morning, of course, it was rice noodles that we had there. They didn’t taste as flavorful as the ‘Hunan rice noodles’ near my school, but they had plenty of ‘mazi’ (a type of topping).’

However, despite the overwhelmingly high proportion of definite objects – a pattern broadly consistent with theoretical accounts describing PC objects as typically definite and uniquely identifiable in discourse – the presence of a non-negligible subset of indefinite or non-referential NP2s complicates this picture. Rather than supporting an absolute definiteness requirement, our data suggest that object definiteness in OFCs constitutes a strong but not obligatory tendency, subject to verb-specific and discourse-pragmatic variation. Notably, a smaller group of verbs, including *chàng* (SD = 69.7%), *mài* (69.2%), *xiě* (69.7%), and *ná* (69.9%), displays more balanced distributions across referential categories, indicating that, alongside definite objects, a substantial number of NP2s are indefinite or non-referential. This is particularly evident where the object is realized as a numeral construction, as in (13), where the verb *ná* introduces a quantity-denoting NP2 that refers to portions of money rather than a specific, identifiable entity:

- 13 那边的条件是：两个人一百元，伙食在外。我一直是拿的六十，那个人拿四十。
(BCC)

wǒ yīzhí shì ná de liùshí
1SG always SHI take DE sixty

‘The conditions over there were: two people, 100 yuan, with meals eaten outside. It was 60 that I always took, and the other person took 40.’

Numeral constructions are attested across nearly all verbs analyzed, albeit representing only a small proportion of NP2s. More striking, however, is the relatively high frequency of objects containing question words such as *shénme* 什么 or *nǎ* 哪 both in the meaning of ‘which’. The objects they select are not a specific, identifiable referent within the discourse; rather, they function to query or highlight a

class of possible entities without committing to a particular one, thereby designating non-referential entities. As Li and Thompson (1989) note, such interrogative pronouns stand for an NP that is unknown to at least one of the interlocutors, a type of object whose existence is presupposed but whose identity is not uniquely specified, in Karttunen's (1977) terms. Examples (14) and (15) illustrate this behavior. In (14), the verb *chàng* takes *nǎ yī-chū* 哪一出 (which act) as its object, leaving the referent underspecified, as the speaker does not know the exact act being performed. Similarly, in (15), *xiě* selects *shénme hútu shī* 什么糊涂诗 (which foolish poems), whereas again *zuò* selects *shénme hútu shì* 什么糊涂事 (which foolish deeds), with both cases referring to objects that are not uniquely identifiable in the discourse.

- 14 其实, 车云菌刚拿到该文时也是一头雾水, 不知魏建军这次又是唱的哪一出?
(zhTenTen)

bù	zhī	Wèi	Jiànjūn	zhè-cì	yòu	shì	chàng
NEG	know	Wei	Jianjun	this-time	again	SHI	sing
de	nǎ	yī-chū					
DE	which	one-CLF					

'Actually, when Cheyunjun first received the article, they were completely baffled and didn't know which was the act that Wei Jianjun was performing this time.'

- 15 只怜云外客, 泪里画红笺.....我那是写的什么糊涂诗, 做的什么糊涂事啊!
(zhTenTen)

wǒ	nà	shì	xiě	de	shénme	hútu	shī	zuò	de	shénme
1SG	there	SHI	write	DE	which	foolish	poem	do	DE	which
	hútu	shì	a							
	foolish	deed	SFP							

'I can only pity the guest from beyond the clouds, painting red notes with tears... What foolish poems was it that I wrote there, what foolish deeds was it that I did!'

Figure 3 visualizes the variability of referentiality in OFCs. While definite objects clearly predominate, a notable proportion of NP2s are found to be indefinite or non-referential, revealing that object behavior in these constructions is more flexible than often assumed and therefore challenging prevailing theoretical accounts that treat PCs' objects as uniformly definite.

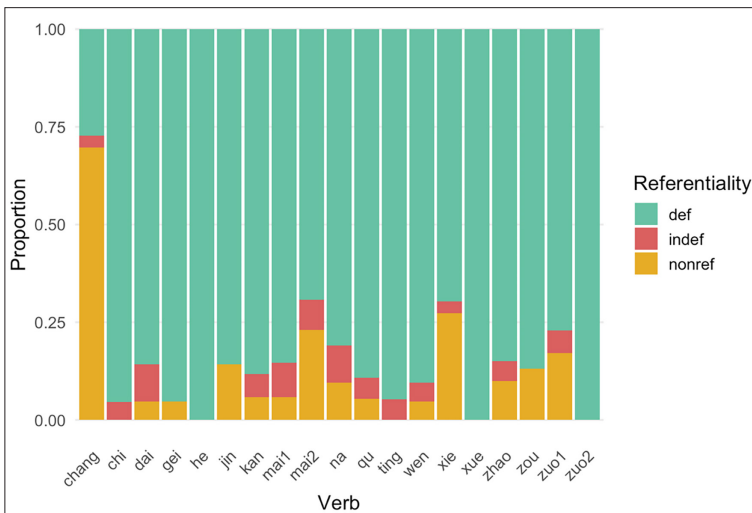


Figure 3 Distribution of NP2 referentiality in OFCs per verb. Chart created by the Authors

5.2.2 Subject Types in OFCs

While previous research has mainly focused on objects as the most interesting argument in PCs, less attention has been devoted to the semantic and referential characteristics of subjects. A series of ANOVAs reveals that concreteness varies significantly across verbs ($F = 3.579$, $p < 1.26e-06$), whereas NP1 animacy and referentiality remain comparatively stable ($F = 1.456$, $p = .169$; $F = 1.357$, $p = .148$, respectively).

In terms of animacy, many verbs display a high degree of internal homogeneity, reflecting the canonical role of the subject as an agentive participant. Although a few verbs – such as *chàng*, *kàn*, *xiě*, and *zhǎo* – show some internal variation ($SD > 0$), the majority consistently select highly animate NP1s. Verbs such as *hē*, *mǎi*, *mài*, *ná*, *qù*, *tīng*, and *zuò2* show no internal variation at all ($SD = 0$), invariably taking human subjects, as in (16).

- 16 如果你是买的新车, 在当地已买购置税就必须要在购买附加税所在地上牌。(zhTenTen)

rúguǒ nǐ shì mǎi de xīn chē
if 2SG SHI buy DE new car

‘If it is a new car that you have bought, and you have already paid the purchase and additional taxes locally, you must register the vehicle in the place where the taxes were paid.’

Overall, 92% of all overt subjects are personal pronouns referring to human participants. Only a small number of tokens involve non-pronominal subjects, typically proper nouns or other NPs, such as *wǒ jiā bǎobao* 我家宝宝 (my baby) in (17). Even in these cases, however, NP1 still denotes human participants.

- 17 我家宝宝一直都是喝的贝因美冠军宝贝奶粉, 很好的。(BCC)

<i>wǒ</i>	<i>jiā</i>	<i>bǎobao</i>	<i>yízhí</i>	<i>dōu</i>	<i>shì</i>	<i>hē</i>	<i>de</i>	<i>Bèiyīnměi</i>
1sg	family	baby	always	all	SHI	drink	DE	Beingmate
<i>Guànjūn</i>	<i>Bǎobèi</i>	<i>Nǎifěn</i>						
Champion	Baby	Formula						

'It is Beingmate Champion Baby Formula that my baby has always been drinking. It's really good.'

The tendency for an animate, actor-like NP to appear in subject position in OFCs can be attributed to its greater topic-worthiness. Animate elements are more likely to emerge as topics because they typically carry out actions, making them central to what speakers choose to comment on and suggesting that agentivity plays a crucial role in representing the 'who' in 'who did what' (Chafe 1976; García, Primus, Himmelmann 2018; Hopper, Thompson 1984). Nevertheless, a small subset of cases includes subjects with lower animacy, as in (18). In this example, the NP *zīběnzhǔyì guójiā* 资本主义国家 (capitalist countries) does not exhibit the same level of animacy as the subject in (17), yet it gains agentive status metonymically, with the institution standing for the people acting on its behalf.

- 18 社会主义医疗保障体系远远好于资本主义, 资本主义国家现在做得比较好的都是学的社会主义。(zhTenTen)

<i>zīběnzhǔyì</i>	<i>guójiā</i>	<i>dōu</i>	<i>shì</i>	<i>xué</i>	<i>de</i>	<i>shèhuìzhǔyì</i>
capitalism	country	all	SHI	learn	DE	socialism

'The socialist medical security system is far superior to the capitalist one. It is socialism that the capitalist countries now performing relatively well have learned from.'

Concreteness, by contrast, shows much greater variability both within and across verbs. Only *mǎi* approaches internal homogeneity (SD < .20), while the remaining verbs select NP1s ranging from medium to highly concrete referents. Since most subjects are personal pronouns, their concreteness tends – by convention (see Brysbaert, Warriner, Kuperman 2014) – to be mid-level rather than fully concrete, as in (19), where the subject *tāmen liǎ* 他们俩 (the two of them) is clearly less concrete than the subject in (18).

- 19 他们俩一个星期全是吃的榆树叶, 连红薯、胡萝卜都没有舍得吃上一口。
(BCC)
tāmen liǎ yī-ge xīngqī quán shì chī de yúshù yè
3PL two one-CLF week entire SHI eat DE elm leaf
'It was only elm leaves that the two of them ate for a whole week, not even
daring to take a bite of sweet potatoes or carrots.'

Turning to referentiality, all NPs in the sample are uniformly definite, regardless of the verb involved. This complete stability contrasts with the variation observed for concreteness but parallels the consistency found for animacy. Just as animate, agentive participants tend to occupy subject position because speakers naturally refer to human agents performing actions, definite NPs are more likely to occur in topic position (Xu 1995; Zhang 2016; Morbiato 2018). This tendency reflects the fact that definite referents encode given information – material assumed to be known, accessible, or recoverable for both interlocutors. This pattern emerges not only in the high frequency of personal pronouns but also in the substantial number of subjects expressed as proper names, as in (20). Here, the subject *Lǎo Ān* 老安 (Lao An) is a definite and identifiable human participant, familiar to both speaker and hearer.

- 20 这次京九通牢采访, 老安是走的人民铁道记者的“后门”。(BCC)
Lǎo.Ān shì zǒu de Rénmín Tiědào bàojìzhě de hòumén
Lao.An SHI walk DE People Railway reporter SP back.door
'For this Jingjiu Tonglao interview, it was the 'back door' of a People's Railway reporter that Lao An went through.'

5.3 Correlation Between Verbs and Focus Types in OFCs

This section explores whether focus type in OFCs is associated with specific verbs or whether it is largely determined by the construction itself, regardless of the verb. To address this, we first tested whether corpus membership affects focus distribution. A logistic regression model was fitted with focus as the dependent variable and both verb and corpus as predictors, estimating the probability of each focus type while controlling for frequency differences. An analysis of deviance revealed that corpus has not significant effect ($Df = 1$, Deviance = 0.003, $p = .959$), indicating that focus patterns are consistent across the zhTenTen and BCC and suggesting a stable, homogeneous behavior of OFCs with respect to this variable.

By contrast, verb as a variable emerged as a highly significant predictor of focus ($Df = 18$, Deviance = 143.27, $p < 2e-16$), showing that different verbs vary systematically in their likelihood of

co-occurring with a particular focus type. Logistic regression quantifies these differences, revealing clear patterns: verbs such as *hē*, *tīng*, and *chī* display moderate probabilities of also co-occurring with contrastive focus (~25-35%), whereas verbs like *chàng* and *zǒu* have a strong to absolute preference for non-contrastive focus (~96-100%). Binomial tests further confirm these patterns, showing that for most verbs, non-contrastive focus is statistically dominant, with the exceptions of *chī* and *tīng*, where neither focus type reaches significance. Figure 4 illustrates these results, with asterisks (*) marking verbs whose dominant focus type is statistically significant.

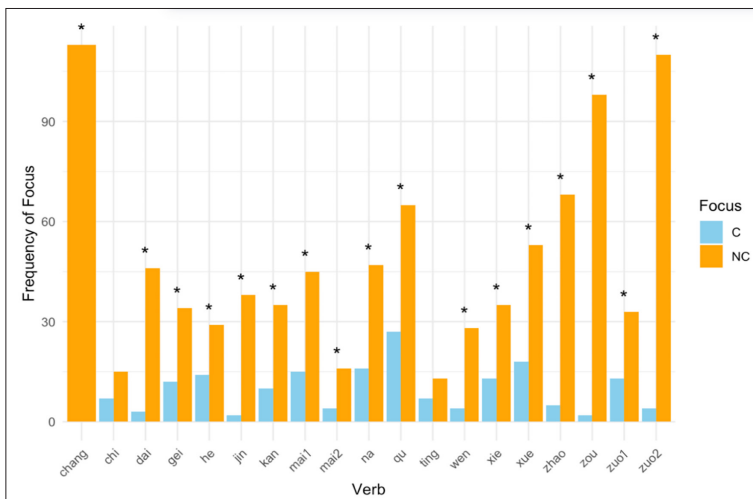


Figure 4 Distribution of focus type in OFCs by verb with statistical significance.
Chart created by the Authors

5.4 Focus Interpretation and Discourse Contexts of OFCs

The qualitative analysis enables a precise identification of the discourse contexts in which OFCs are used to realize focus, either in non-contrastive or contrastive interpretations. Non-contrastive interpretations of focus – by far the most frequently attested across all verbs occurring in OFCs – can be further differentiated into three main discourse contexts: (i) narrative-descriptive contexts, in which OFCs present information about completed events without evoking alternative sets; (ii) wh-question contexts, where OFCs provide information about a previously established event without any contrastive intent; and (iii) anaphoric contexts, in which OFCs reintroduce a previously mentioned referent, promoting it from the background to the foreground and ensuring discourse continuity.

The narrative-descriptive context is observed across all verb classes. It typically occurs when speakers recount habitual practices, travel experiences, childhood stories, or other memories from the past, specifying the object involved in the realized event. In this context, the object is in focus. This use does not imply contrast with alternative elements; rather, it serves to explicitly identify the object as the salient element of the event, making clear what was actually being done. In this way, OFCs bring the object into attention and support a detailed, scene-building description of past experiences and practices.

For example, in (21), the object *zǎoyóuwán* 藻油丸 (algae oil) is in focus, as it specifies which capsules were taken at that particular moment. The discourse context is narrative, situating the event within a past habitual setting. In this context, OFCs present information about the object involved in the event without implying opposition to alternative items. This use introduces information that is new to the listener and presents the event in a detailed and explicit way, without any intention of creating contrast or implying that the focused element is unexpected.

- 21 我一般在早晨吃饭后服用, 我当时是吃的香港美赞臣的DHA藻油丸, 吃起来很方便。(zhTenTen)

wǒ dāngshí shì chī de Xiānggǎng Měizhànrén de
1SG then SHI eat DE Hong Kong Mead Johnson SP
DHA zǎoyóu wán
DHA algae.oil capsules

'I usually take it after breakfast in the morning. At that time, it was DHA algae oil capsules from Mead Johnson Hong Kong that I was taking. They are very convenient to take.'

OFCs can also occur within wh-questions, both direct and indirect, when the object is a wh- element. In such cases, one important function of OFCs is to obtain information about an event that has already occurred, without implying any contrast. For example, in (22), the object is *shénme jiǎnchá* 什么检查 (which test). Here, the doctor is asking purely to gather information; the OFC does not imply contrast with other tests or possibilities, nor does it presuppose anything contrary to the hearer's expectations or prior knowledge. Instead, it serves to establish a general picture of the patient's health, which is necessary for prescribing appropriate care.

- 22 如果您担心复查是阳性, 也可以立刻做一个疗程的正规治疗。您当时是做的什么检查呢? 建议做个TPPA检查, 明确自己的情况。(zhTenTen)

nín dāngshí shì zuò de shénme jiǎnchá ne
2SG then SHI do DE what test SFP

'If you are concerned that the follow-up test might be positive, you can also immediately start a regular treatment course for the child. Which test was it that you had at the time? It is recommended to take a TPPA test to clarify your situation.'

Finally, regarding the anaphoric context, OFCs are used to reintroduce previously mentioned or known elements, allowing the speaker to maintain reference continuity and ensuring that the listener can track entities or events already established in the discourse. This context is particularly frequent with verbs such as *qù* and *xué*.

For example, in (23), the object 4S *diàn* 店 (dealership) following *shì bù shì qù de* 是不是去的 serves to recall an element previously introduced, namely the same dealership:

- 23 去哪里维修保养都无所谓了, 当然咱自己的车, 图个安心, 去4S店也行.....因为 是二手车, 我们不知道以前是不是去的4S店做保养.....(zhTenTen)

wǒmen bù zhīdào yíqián shì bù shì qù de 4S
1SG NEG know before SHI NEG SHI go DE 4S
diàn zuò bǎoyǎng
dealership do maintenance

'It doesn't really matter where you go for maintenance or repairs. Of course, for our own car, just for peace of mind, going to a 4S dealership is also fine. Since it's a used car, we don't know whether it was a 4S dealership that one went to for maintenance.'

Here, the OFC brings a previously mentioned element from the background to the foreground. The construction does not imply contrast; rather, it redirects the listener's attention to this specific entity, emphasizing precisely which workshop the maintenance might have occurred at. In other words, the OFC highlights a known element from the prior context, ensuring that the listener connects the current statement to what has already been introduced, thereby maintaining discourse coherence.

Regarding contrastive focus, both explicit and implicit contrastive cases are observed in OFCs. Explicit contrastive focus is rare in the corpus – about 25% of contrastive OFCs – and mainly occurs with verbs of possession change and instrumental locomotion. In some cases, these OFCs can also convey corrective contrast. For example, in (24), the first OFC marks a contrastive focus that rejects the hearer's assumption that the acquired item was a yogurt maker. The second OFC in the same sentence then functions as a corrective

focus, specifying the actual item acquired and setting it in opposition to the previously rejected, incorrect assumption.

- 24 哈哈, 是我买过的店铺, 不过不是买的酸奶机, 是买的一套托玛琳的护腰三件套。 (zhTenTen)
- | | | | | | | | | |
|---------------|-----------------|------------|---------------|----------------|------------------|------------|------------|-----------|
| <i>bùguò</i> | <i>bù</i> | <i>shì</i> | <i>mǎi</i> | <i>de</i> | <i>suānnǎijī</i> | <i>shì</i> | <i>mǎi</i> | <i>de</i> |
| but | NEG | SHI | buy | DE | yogurt-maker | SHI | buy | DE |
| <i>yī-tào</i> | <i>tuǒmǎlín</i> | <i>de</i> | <i>hù yāo</i> | <i>sān-jàn</i> | <i>tào</i> | | | |
| one-CLF | tourmalines | SP | lumbar. | three-CLF | set | | | |
| | | | support | | | | | |
- ‘Haha, this is the shop I bought from before, but it wasn’t the yogurt maker that I bought; it was a three-piece tourmaline lumbar support set that I bought.’

Implicit contrastive focus, on the other hand, accounts for the majority of the contrastive focus cases observed in the analyzed OFCs and is especially common with verbs of ingestion or carrying. For example, in (25), the sentence conveys an implicit contrast – not between two types of food as objects, but between expectations and reality regarding the effect of the food. The object *gǒuliáng* 狗粮 (dog food) is in focus using the OFC construction, drawing attention to the specific type of food the puppy has been eating. This illustrates an implicit contrast, which can emerge both from an imaginable set of alternatives and from the pragmatic and discourse context: one would expect that switching to proper dog food would eliminate bad breath, yet the puppy’s breath still smells unpleasant despite the dietary improvement. The focus does not juxtapose two different objects directly; rather, it emphasizes the salient object central to the event while allowing the listener to perceive the unexpected outcome. This contrast is further reinforced by the temporal coordinates, contrasting a ‘before’, when the puppy had just arrived at home, with a more recent time.

- 25 狗狗买回来的时候嘴里就特别的臭, 那时候才两月, 最近都是吃的狗粮, 可是嘴里还是臭的。 (zhTenTen)
- | | | | | | | | |
|---------------|------------|-------------|------------|-----------|-----------------|--------------|---------------|
| <i>zuìjìn</i> | <i>dōu</i> | <i>shì</i> | <i>chī</i> | <i>de</i> | <i>gǒuliáng</i> | <i>kěshì</i> | <i>zuǐ-lǐ</i> |
| lately | all | SHI | eat | DE | dog.food | but | mouth-inside |
| <i>hái</i> | <i>shì</i> | <i>chòu</i> | <i>de</i> | | | | |
| still | SHI | stinky | SFP | | | | |
- ‘When we bought the puppy, its mouth already smelled really bad, and it was only two months old at the time. Lately, it’s only dog food that it’s been eating, but its mouth still smells bad.’

6 General Discussion and Conclusions

The combined statistical and qualitative analyses presented in this study offer a clearer understanding of a relatively understudied subtype of Chinese PCs, i.e., the OFCs. The distributional patterns of the 20 most frequent Chinese dynamic verbs show that, with the exception of *lái*, most verbs do occur in this construction. Notably, verbs such as *zǒu*, *mǎi*, *zuò*², *qù*, and *kàn* are particularly frequent, strongly mirroring the discourse environments in which OFCs typically arise. Text type analysis further indicates that the OFC is particularly common in social media, a genre that has recently attracted growing attention in linguistic research due to its rich and creative linguistic repertoire and its proximity to spoken language (Zhang, Ren 2022). As Lomborg (2011) notes, content generated by ordinary users reflects everyday personal lives, so the language employed closely resembles informal, conversational speech. This context therefore likely favors verbs describing dynamic, agentive, and concrete actions, which aligns with the high frequency of the verbs identified in the corpora.

The analysis of the construction's semantic arguments further reveals systematic asymmetries between subjects and objects. Subjects are consistently animate and predominantly human, whereas objects are almost uniformly inanimate. This configuration aligns with well-established cross-linguistic tendencies in role prototypicality, according to which animate entities are preferentially construed as actors, and inanimate entities as patients (García, Primus, Himmelmann 2018; Primus 2011). Prior research on Chinese, including Xu (1995), Philipp et al. (2008), and Wu, Kaiser and Andersen (2011), has demonstrated that animacy exerts strong effects on word order, whereby more animate NPs tend to occupy subject position and display higher topicality, while inanimate NPs cluster in object position, due to the 'animate first' preference. This general tendency also helps explain why NP1s in OFCs are almost exclusively definite and discourse-identifiable, given the strong association between sentence-initial position, topichood, and given information (Morbiato 2018).

NP2s, by contrast, show more variable referential behavior. While objects in Chinese may be definite or indefinite (Chen 2004), the literature on PCs typically assumes that the object in these constructions must be definite. The present data, however, complicate this assumption. Although many NP2s are indeed definite, a substantial portion are found to be indefinite or even non-referential. These results indicate that the object of an OFC – although canonically described as uniquely identifiable – can in fact display verb-dependent flexibility, sometimes appearing with numeral-classifier constructions or even question-word forms.

Regarding focus interpretation, the analyses indicate that OFCs predominantly encode non-contrastive focus, with only a small set of verbs deviating from this tendency. This pattern suggests that focus behavior is largely construction-driven rather than verb-driven.

Non-contrastive OFCs can be observed in three main discourse contexts: (i) narrative-descriptive contexts, in which they present information about specific objects in relation to completed past events, typically in memories, travels, or routine experiences; (ii) wh-question contexts, where the object is unknown and OFCs provide purely informational content without contrastive implications; and (iii) anaphoric contexts, in which they reintroduce a previously mentioned object and bring it back into the discourse foreground, maintaining reference continuity and allowing the listener to track discourse-relevant entities.

By comparison, contrastive OFCs are infrequent, contrary to previous theoretical claims (see Jing-Schmidt 2017). The most common type is implicit contrast, arising from contextual inference rather than explicit opposition. Explicit contrast, including corrective focus, is rarer and tends to cluster with verbs of transfer, ingestion, and locomotion, where clear alternatives are expressed or corrected.

This study has several limitations, which also suggest directions for future research. First, intended as a preliminary corpus-based overview of Chinese OFCs, the analysis provides only an initial account of focus interpretation. Future research could develop a more fine-grained taxonomy of object focus realization in OFCs and its discourse contexts, mapping the full range of contrastive and non-contrastive focus types across these contexts, and extending the analysis to other types of PCs. Second, since cleft interpretation is often influenced by prosody, particularly pitch accent and stress patterns (Zimmermann 2008), future studies should incorporate spoken data to examine how prosodic prominence interacts with focus interpretation. Elicited judgments and interviews could further clarify speakers' communicative intentions, for instance whether object focus is intended as contrastive or whether contrastive readings emerge from discourse context. Finally, the findings suggest that reference grammars and pedagogical materials would benefit from a more accurate description of OFCs, which could help learners and teachers better understand their use in authentic discourse beyond prescriptive accounts.

Bibliography

- Bajčetić, L.; Declerck, T. (2022). Using Wiktionary to Create Specialized Lexical Resources and Datasets. *Proceedings of the 13th Conference on Language Resources and Evaluation (Marseilles, 20-25 June 2022)*. Marseilles: European Language Resources Association, 3457-60.
- Berretta, M. (1994). "Ordini marcati dei costituenti di frase in italiano. La frase scissa". *Vox Romanica*, 53, 79-105.
- Biber, D.; Johansson, S.; Leech, G.; Conrad, S.; Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Bolinger, D. (1961). "Contrastive Accent and Contrastive Stress". *Language*, 37, 83-96. <https://doi.org/10.1017/S0025100300002061>.
- Brysbaert, M.; Warriner, A.B.; Kuperman, V. (2014). "Concreteness Ratings for 40 Thousand Generally Known English Word Lemmas". *Behavior Research Methods*, 46, 904-11. <https://doi.org/10.3758/s13428-013-0403-5>.
- Callies, M. (2009). *Information Highlighting in Advanced Learner English. The Syntax-Pragmatics Interface in Second Language Acquisition*. Amsterdam: John Benjamins.
- Chafe, W. (1976). "Givenness, Contrastiveness, Definiteness, Subjects, Topic and Point of View". Li, C. (ed.), *Subject and Topic*. New York: Academic Press, 25-56.
- Chang, L.-L.; Chen, K.-J.; Huang, C.-R. (2000). "A Lexical-Semantic Analysis of Mandarin Chinese Verbs: Representation and Methodology". *Computational Linguistics and Chinese Language Processing*, 5(1), 1-18.
- Chen, P. (2004). "Identifiability and Definiteness in Chinese". *Linguistics*, 42(6), 1129-84. <https://doi.org/10.1515/ling.2004.42.6.1129>.
- Chen, P. (2009). "Aspects of Referentiality". *Journal of Pragmatics*, 41(8), 1657-74. <https://doi.org/10.1016/j.pragma.2009.03.002>.
- Cheng, L. (2008). "Deconstructing the *shi...de* Construction". *The Linguistic Review*, 25(3-4), 235-66.
- Chin, C.O. (2010). "The GO-Type and the GIVE-Type Indirect Object Markers in the Conghua Dialect". Coblin, W.S.; Yue, O.A. (eds), *Studies in Honor of Jerry Norman*. Hong Kong: The Ng Tor-tai Chinese Language Research Centre, Institute of Chinese Studies, The Chinese University of Hong Kong, 245-75.
- Comrie, B. (1989). *Language Universals and Linguistic Typology*. Chicago: University of Chicago Press.
- Cui, S.; Sung, K. (2021). *A Reference Grammar for Teaching Chinese. Syntax and Discourse*. Singapore: Springer; Peking University Press.
- Cummins, G.M. III (1998). "Definiteness in Czech". *Studies in Language*, 22(3), 567-96.
- Declerck, R. (1984). "The Pragmatics of IT-Clefts and WH-Clefts". *Lingua*, 64, 251-89.
- Declerck, R. (1988). *Studies on Copular Sentences, Clefts and Pseudo-Clefts*. Berlin; New York: De Gruyter.
- Dik, S. (1997). *The Theory of Functional Grammar. Part 1, The Structure of the Clause*. Berlin: Mouton de Gruyter.
- Garassino, D. (2016). "Using Cleft Sentences in Italian and English. A Multifactorial Analysis". De Cesare, A.M.; Garassino, D. (eds), *Current Issues in Italian, Romance and Germanic Non-Canonical Word Orders. Syntax-Information Structure-Discourse Organization*. Frankfurt: Peter Lang, 181-204. <https://doi.org/10.3726/978-3-653-05662-4/19>.
- García, M.G.; Primus, B.; Himmelmann, N.P. (2018). "Shifting from Animacy to Agentivity". *Theoretical Linguistics*, 44, 25-39. <https://doi.org/10.1515/tl-2018-0002>.

- Givón, T. (2001). *Syntax. An Introduction*, vol. 1. 2nd ed. Amsterdam: John Benjamins.
- Halliday, M.A.K. (1967). "Notes on Transitivity and Theme in English". *Journal of Linguistics*, 3, 37-81. <https://doi.org/10.1017/S0022226700012949>.
- Han Mei 韩梅 (2005). "'Shi...de' jude jufa yuyi fenxi" "是...的"句的句法语义分析 (A Semantic-Syntactic Analysis of the 'shi...de' Pattern). *Dongjiang Journal*, 22(2), 97-103.
- Hartmann, K.; Veenstra, T. (2013). "Introduction". Hartmann, K.; Veenstra, T. (eds), *Cleft Structure*. Amsterdam: John Benjamins, 1-32.
- Hole, D. (2011). "The Deconstruction of Chinese shì...de Clefts Revisited". *Lingua*, 121, 1707-33. <https://doi.org/10.1016/j.lingua.2011.07.004>.
- Hole, D.; Zimmermann, M. (2013). "Cleft Partitioning in Japanese, Burmese, and Chinese". Hartmann, K.; Veenstra, T. (eds), *Cleft Structures*. Amsterdam: John Benjamins, 285-318. <https://doi.org/10.1075/la.208.11hol>.
- Hopper, P.J.; Thompson, S.A. (1984). "The Discourse Basis for Lexical Categories in Universal Grammar". *Language*, 60(4), 703-52.
- Iurato, A. (2024a). "Exploring the Pragmalinguistic Knowledge of the 是 shì...的 de Cleft Construction In L1 Italian Learners' L2 Chinese: Triangulating Corpus and Experimental Data". Kecskes, I.; Zhang, H. (eds), *Chinese as a Second Language from Different Angles*. Leiden: Brill, 76-110.
- Iurato, A. (2024b). "Design and Collection of the Bimodal Italian Learner Corpus Of L2 Chinese (BILCC)". Ackerley, K.; Castello, E. (eds), *Continuing Learner Corpus Research: Challenges and Opportunities*. Louvain-la-Neuve: Presses universitaires de Louvain, 21-54.
- Iurato, A. (2024c). *The Acquisition of the Chinese 是 shì...的 de Cleft Construction: Triangulating Corpus and Experimental Data* [PhD Dissertation]. Venice: Ca' Foscari University of Venice; Bremen: University of Bremen.
- Iurato, A. (2026). "The Acquisition of Chinese shì...de Proper Clefts by L1 Italian Learners: A Corpus-Based Approach". *Global Chinese*, 12(1), 111-39. <https://doi.org/10.1515/glochi-2025-0005>.
- Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge (MA): MIT Press.
- Jiang, L.J. (2015). "Marking (in) Definiteness in Classifier Languages". *Bulletin of Chinese Linguistics*, 8(2), 319-43. <https://doi.org/10.1163/2405478X-00802010>.
- Jing-Schmidt, Z. (2017). "Grammatical Construction and Chinese Discourse". Shei, C. (ed.), *The Routledge Handbook of Chinese Discourse Analysis*. London; New York: Routledge, 209-30.
- Karttunen, L. (1977). "Syntax and Semantics of Questions". *Linguistics and Philosophy*, 1(1), 3-44. <https://doi.org/10.1075/sl.22.3.03cum>.
- Kenesei, I. (2006). "Focus as Identification". Molnár, V.; Winkler, S. (eds), *The Architecture of Focus*. Berlin: Mouton de Gruyter, 137-68.
- Kiss, K.É. (1998). "Identificational Focus Versus Informational Focus". *Language*, 74(2), 245-73.
- König, E. (1991). *The Meaning of Focus Particles: A Comparative Perspective*. London: Routledge.
- Korzen, I. (2014). "Cleft Sentences. Italian-Danish in Contrast". De Cesare, A.M. (ed.), *Frequency, Forms and Functions of Cleft Constructions in Romance and Germanic: Contrastive, Corpus-Based Studies*. Berlin; München; Boston: De Gruyter, 217-75.
- Krifka, M. (1999). "Additive Particles Under Stress". Strollovitch, D.; Lawson, A. (eds), *Proceedings of the 8th Semantics and Linguistic Theory Conference* (Cambridge, MA, 8-10 May 1998). Cornell: CLC Publications, 111-28.

- Lambrecht, K. (1994). *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.
- Lambrecht, K. (2001). "A Framework for the Analysis of Cleft Constructions". *Linguistics*, 39(3), 463-516.
- LaPolla, R.J. (1990). *Grammatical Relations in Chinese: Synchronic and Diachronic Considerations* [PhD Dissertation]. Berkeley: University of California.
- Levin, B. (1993). *English Verb Classes and Alternations*. Chicago: University of Chicago Press.
- Li, C.N.; Thompson, S.A. (1989). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.
- Li, K. (2008). "Contrastive Focus Structure in Mandarin Chinese". Marjorie, K.M.; Hana, K. (eds), *Proceedings of the 20th North American Conference of Chinese Linguistics* (Columbus, OH, 25-27 April 2008), vol. 2. Columbus: The Ohio State University Press, 759-74.
- Liu, M. (2015). Tense and Aspect in Mandarin Chinese. Wang, W.S.-Y.; Sun, C. (eds), *The Oxford Handbook of Chinese Linguistics*. Oxford: Oxford University Press, 274-89.
- Liu, Y. (2023). *Exhaustivity, Contrastivity, and the Semantics of Mandarin Cleft-Related Structures*. Oxon; New York: Routledge.
- Locatelli, L. (2026). *Chinese Passive Constructions in Environmental Discourse. A Corpus-Assisted Functional Study from an Ecolinguistic Perspective* [PhD Dissertation]. Venice: Ca' Foscari University of Venice.
- Lomborg, S. (2011). "Social Media as Communicative Genres". *MedieKultur: Journal of Media and Communication Research*, 27(51), 55-71. <https://doi.org/10.7146/mediekultur.v27i51.4012>.
- Long, H.-P. (2013). "On the Formation of Mandarin VdeO Focus Clefts". *Acta Linguistica Hungarica*, 60(4), 409-56.
- Luo, C. (2009). *Cleftability in Language*. Wuhan: Wuhan University Press.
- Mai, Z.; Yuan, B. (2016). "Uneven Reassembly of Tense, Telicity and Discourse Features in L2 Acquisition of the Chinese *shì...de* Cleft Construction by Adult English Speakers". *Second Language Research*, 32(2), 247-76.
- Miller, P.H. (2006). "Focus in the Languages of Europe". Bernini, G.; Schwartz, M. (eds), *Pragmatic Organization of Discourse in the Languages of Europe*. Berlin; New York: De Gruyter, 121-214.
- Morbiato, A. (2018). *Word Order and Sentence Structure in Mandarin Chinese: New Perspectives* [PhD Dissertation]. Venice: Ca' Foscari University of Venice.
- Morbiato, A.; Arcodia, G.F.; Basciano, B. (2020). "Topic and Subject in Chinese and in the Languages of Europe: Comparative Remarks and Implications for Chinese as a Second/Foreign Language Teaching". *Chinese as a Second Language Research*, 9(1), 31-66. <https://doi.org/10.1515/caslar-2020-0002>.
- Nikolaeva, I. (2001). "Secondary Topic as a Relation in Information Structure". *Linguistics*, 39, 1-49.
- Paul, W. (2005). "Low IP Area and Left Periphery in Mandarin Chinese". *Recherches linguistiques de Vincennes*, 33, 111-34.
- Paul, W. (2015). *New Perspectives on Chinese Syntax*. Berlin; Boston: De Gruyter.
- Paul, W.; Whitman, J. (2008). "Shì...de Focus Clefts in Mandarin Chinese". *The Linguistic Review*, 25(3-4), 413-51.
- Philipp, M.; Bornkessel-Schlesewsky, I.; Bisang, W.; Schlewsky, M. (2008). "The Role of Animacy in the Real Time Comprehension of Mandarin Chinese: Evidence from Auditory Event-Related Brain Potentials". *Brain and Language*, 105(2), 112-33. <https://doi.org/10.1016/j.bandl.2007.09.005>.

- Primus, B. (2011). "Animacy, Generalized Semantic Roles, and Differential Object Marking". Lamers, M.; de Swart, P. (eds), *Case, Word Order and Prominence: Interacting Cues in Language Production and Comprehension*. Dordrecht: Springer Netherlands, 65-90. https://doi.org/10.1007/978-94-007-1463-2_4.
- Prince, E. (1981). "Toward a Taxonomy of Given/New Information". Cole, P. (ed.), *Radical Pragmatics*. New York: Academic Press, 223-55.
- Reeve, M. (2013). "The Cleft Pronoun and Cleft Clause in English". Hartmann, K.; Veenstra, T. (eds), *Cleft Structures*. John Benjamins, 165-86.
- Repp, S. (2010). "Defining 'Contrast' as an Information-Structural Notion in Grammar". *Lingua*, 120(6), 1333-45. <https://doi.org/10.1016/j.lingua.2009.04.006>.
- Rochemont, R. (1986). *Focus in Generative Grammar*. Amsterdam: John Benjamins.
- Rooth, M. (1992). "A Theory of Focus Interpretation". *Natural Language Semantics*, 1(1), 75-116.
- Ross, C. (1983). "On the Functions of Mandarin de". *Journal of Chinese Linguistics*, 11(2), 215-46.
- Shi, D. (1994). "The Nature of Chinese Emphatic Sentences". *Journal of East Asian Linguistics*, 3, 81-100.
- Shyu, S. (2015). s.v. "Shì 是.....de 的 Sentences". Sybesma, R. (ed.), *Encyclopedia of Chinese Language and Linguistics*. http://dx.doi.org/10.1163/2210-7363_ecLL_COM_00000092.
- Simpson, A.; Wu, Z. (2002). "From D to T- Determiner Incorporation and the Creation of Tense". *Journal of East Asian Linguistics*, 11, 169-209.
- Song Yuzhu 宋玉株 (1978). "Guanyu Shi...de jiegou de fenxi: yufa biji yi ze" 关于“是.....的”结构的分析：语法笔记一则 (An Analysis of the 'shi...de' Structure – A Grammar Note). *Tianjin Shiyuan Xuebao*, 39, 75-6.
- Sugimura Hirofumi 杉村博文 (1999). "De' zi jiegou, chengzhi yu fenlei" 的“字结构，承旨与分类 (The Structure, the Origin and Classification of the Character 'de'). Jiang Lansheng 江蓝生; Hou Jing 侯精 (eds), *Hanyu xianzhuang yu lishi de yanjiu* 汉语现状与历史的研究. Beijing: China Social Sciences Press, 25-32.
- Vallduví, E.; Vilkkuna, M. (1998). "On Rheme and Contrast". *Syntax and Semantics*, 29, 79-108.
- Wu, F.; Kaiser, E.; Andersen, E. (2011). "Animacy Effects in Chinese Relative Clause Processing". *Language and Cognitive Processes*, 27(10), 1489-1524. <https://doi.org/10.1080/01690965.2011.614423>.
- Xu, L. (1995). "Definiteness Effects on Chinese Word Order". *Cahiers de linguistique-Asie orientale*, 24(1), 29-48. <https://doi.org/10.3406/clao.1995.1465>.
- Xu, Y. (2014). "A Corpus-Based Functional Study of shi...de Constructions". *Chinese Language and Discourse*, 5(2), 146-84. <https://doi.org/10.1075/cld.5.2.02xu>.
- Zhang, D. (2016). "Definiteness and the Meaning Reconstruction in English-Chinese Translation". *Babel: Revue internationale de la traduction/International Journal of Translation*, 62(3), 456-69. <https://doi.org/10.1075/babel.62.3.05zha>.
- Zhang, Y.; Ren, W. (2022). "This is so skrrrrr'-Creative Translanguaging by Chinese Micro-Blogging Users". *International Journal of Multilingualism*, 19(3), 289-304. <https://doi.org/10.1080/14790718.2020.1753746>.
- Zhao Shuhua 赵淑华 (1979). "Guanyu 'shi...de' jushi" 关于“是...的”句式 (On 'shi...de' Sentences). *Yuyan Jiaoxue yu Yanjiu*, 1, 57-66.
- Zimmermann, M. (2008). "Contrastive Focus and Emphasis". *Acta Linguistica Hungarica*, 347-60. <https://doi.org/10.1556/aling.55.2008.3-4.9>.