Chapter 23
Linking Comprehension and Production: Frequency Distribution of Chinese Relative Clauses in the Sinica Treebank

Chien-Jer Charles Lin and Hai Hu

Abstract This chapter presents the distribution of Chinese relative clauses in the Sinica Treebank (Chen et al., Sinica corpus: Design methodology for balanced corpora, 1996; Huang et al., *Mandarin Chinese words and parts of speech: A corpus-based study*, Taylor & Francis, 2017). We extracted 3081 relative clauses from the treebank and classified the relative clauses into six types, including gapless relative clauses, possessive relative clauses, descriptive relative clauses, passive relative clauses, subject relative clauses, and object relative clauses. Each type of relative clause will be discussed regarding the length and syntactic complexity of prenominal clauses, the length and animacy/humanness of head nouns, the part-of-speech categories of embedded verbs, and the position of complex noun phrases in matrix clauses. The issues of the classifier phrase position in relation to relative clauses, the use of *suo* in object relative clauses, and cases where the head nouns are omitted will also be discussed. Based on the corpus distributions, we consider the implications for the comprehension of Chinese relative clauses.

Keywords Treebank · Chinese relative clauses · Sentence processing · Production-distribution-comprehension model · Corpus linguistics

C.-J. C. Lin (✉)
Department of East Asian Languages and Cultures, Indiana University Bloomington, Bloomington, IN, USA
e-mail: chiclin@indiana.edu

H. Hu
School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China
e-mail: hu.hai@sjtu.edu.cn

23.1 Introduction: The Corpus's Role in Sentence Processing

An important topic in linguistic research concerns the interface issue, namely, how a language system interacts with computation, expressive content, and articulation. Two dimensions of linguistic processing—language comprehension and language production—are particularly important. Language comprehension revolves around how the mind perceives and interprets linguistic signals, whereas language production entails how the mind generates linguistic codes for articulation. These dual facets of language processing serve as the foundation for the development of various theories concerning language.

While it might appear evident that there should be a connection between language comprehension and language production, the precise nature of this connection remains less clear. A conventional model like the Speech Chain (Denes and Pinson 1993) sees comprehension and production as inseparable facets of the same coin. Language production corresponds to the speaker (i.e., encoding) aspect of the chain while language comprehension corresponds to the listener (i.e., decoding) aspect of the chain. Such models typically assume a symmetrical relation between comprehension and production, with these two aspects linked through shared linguistic representations. Accordingly, if a linguistic expression is difficult to encode, it is also taken to be difficult to decode. The complexity of linguistic representations and language users' experience with language comprehension and language production can both account for the symmetrical processing effects in comprehension and production. Linguistic materials that are more complex are expected to be harder to interpret and produce (Ferreira 1991; Gibson and Warren 2004). Similarly, less frequently encountered/produced expressions are expected to be more demanding to understand (Reali and Christiansen 2007).

The Production-Distribution-Comprehension (PDC) model (MacDonald 2013) represents a significant endeavor to directly bridge the realms of sentence production and sentence comprehension. According to the PDC model, the distributional regularities in corpora provide valuable insights into the mechanisms at play during utterance planning. This involves organizing information based on processing ease, with a tendency to reuse recently employed structures. Distributional regularities can also be used to predict how utterances may unfold (Hale 2001, 2006; Levy 2008). Distributional regularities from corpora therefore serve as an important resource for making inferences about grammar. On the one hand, corpus data can be seen as a snapshot of collective language production, revealing what structures and expressions are favored in a given context. On the other hand, corpus data illuminates the probabilistic underpinnings of grammar based on which parsing decisions are made.

23.2 Processing Relative Clauses

Taking relative clauses (RCs) as an example, a common finding in English is that subject-extracted relative clauses (SRCs) like (1) below are easier to process than object-extracted relative clauses (ORCs) like (2) both for comprehension and for production (Gibson et al. 2005; King and Just 1991; Traxler et al. 2002; see Lin and Bever 2006 and O'Grady 2011 for a typological overview). Multiple factors contrasting SRCs and ORCs can account for the processing advantage of (1) over (2), including, for instance, the shorter distance between the head and the gap in SRCs compared with that in ORCs (Gibson 1998) and the canonical thematic order of Noun-Verb-Noun (NVN) or Agent-Verb-Patient found in SRCs but not in ORCs (Bever 1970; Lin 2014, 2015).

(23.1)	The harpist$_i$ who [GAP$_i$] knows the composer received good reviews.
(23.2)	The harpist$_i$ who the composer knows [GAP$_i$] received good reviews.

The processing advantage of SRCs is predicted based on the formal property of the linguistic material, namely, a shorter filler-gap distance and the canonicity of word orders found in SRCs. Intriguingly, this processing asymmetry is also consistent with the distributional dominance of SRCs in corpora. Roland et al. (2007), for instance, reported that ORCs are less frequent than SRCs in English written corpora. Considering production, distribution, and comprehension, therefore, RCs in English show a rather consistent pattern; that is, SRCs exhibit higher frequency and are generally easier to process compared to ORCs.

The underlying reasons for this correlation, however, remain a subject of debate, given the presence of multiple factors that can make similar predictions. One potential scenario considers production as the foundation for distributional dominance and, consequently, ease of comprehension. In this view, due to factors like locality and word order canonicity, planning the production of an SRC is inherently more straightforward than that of an ORC. Consequently, SRCs tend to appear more frequently in corpora. As language users encounter SRCs more often, they become more adept at both producing and comprehending them, creating a self-reinforcing cycle. Another plausible scenario involves inferring from frequency distribution that SRCs serve a more functional role in discourse than ORCs. Given their higher frequency of use, SRCs are not only easier to produce or reuse but are also more likely to be expected and comprehended by language users. Several other explanations could account for this correlation, but the linked observations in comprehension, production, and corpus distribution have yet to definitively establish the causal relationships among them.

This chapter will report the distributional frequencies of Chinese relative clauses in the Sinica Treebank 3.0 (http://turing.iis.sinica.edu.tw/treesearch/; Chen et al. 1996, 2003) and discuss these distributions in light of their significance in sentence processing. In recent years, researchers have increasingly focused on the processing of head-final relative clauses, where RCs appear before the head nouns they modify.

Chinese, in particular, has garnered attention in sentence processing research. While the basic word order of Chinese is Subject-Verb-Object (SVO) as it is in English, the noun phrase (NP) structure in Chinese is head-final. The embedded clause in a Chinese NP appears before the noun it modifies. Owing to this typological particularity, SRCs and ORCs in Chinese present distinct filler-gap relations than those in English. Specifically, Chinese RCs feature gaps that precede fillers in terms of linear order, and SRCs entail longer dependency distances compared to ORCs as shown in (3-4). Furthermore, ORCs, but not SRCs, adhere to the canonical NVN order in Chinese. These considerations related to locality and word order suggest a processing advantage for ORCs over SRCs, in contrast to the observations in English.

(23.3) [GAP$_i$]認識作曲家的豎琴家$_i$獲得好評。
[GAP$_i$]__renshi__zuoqujia__de__shuqinjia$_i$__huode__haoping
[GAP$_i$]__know__composer__DE__harpist$_i$__win__good.review
The harpist$_i$ who [GAP$_i$] knows the composer received good reviews.
(23.4) 作曲家認識 [GAP$_i$]的豎琴家$_i$獲得好評。
zuoqujia__renshi__[GAP$_i$]__de__shuqinjia$_i$__huode__haoping
composer__know__[GAP$_i$]__DE__harpist$_i$__win__good.review
The harpist$_i$ who the composer knows [GAP$_i$] received good reviews.

Head-final relative clauses like those in Chinese therefore offer an intriguing arena for the various comprehension and production factors that have otherwise been complicated in head-initial RCs. While locality and word order canonicity both predict easier comprehension of SRCs in English, they predict easier comprehension of ORCs in Chinese. Interestingly, the distribution of relative clauses in Chinese corpora does not consistently align with these processing predictions as observed in English. Frequency distributions have quite consistently indicated higher occurrence of SRCs than ORCs in the corpora (e.g., Wu et al. 2011), thus predicting an SRC advantage. In fact, research on Chinese RC processing has yielded mixed results. In terms of comprehension, some studies have reported that SRCs are easier (Chen et al. 2012; Jäger et al. 2015; Lin and Bever 2006), while others have reported that ORCs are easier (Gibson and Wu 2013; Hsiao and Gibson 2003; Lin 2014; Lin and Garnsey 2011; Packard et al. 2011; Qiao et al. 2012; Sung et al. 2016). In terms of RC production, SRCs have been found to take a shorter time to initiate than ORCs (Lin 2013).

The dominance of SRCs in corpora is in line with the SRC advantage in sentence planning (Lin 2013) and in some comprehension studies (Chen et al. 2012; Jäger et al. 2015; Lin and Bever 2006) but in conflict with the ORC advantage in other comprehension studies (Gibson and Wu 2013; Hsiao and Gibson 2003; Lin 2014; Lin and Garnsey 2011; Packard et al. 2011; Qiao et al. 2012; Sung et al. 2016). In light of this, our study aims to delve deeper into the distributions of Chinese RCs while considering their relevance to critical issues in RC processing. Subsequent sections will dissect the corpus data extracted from the Sinica Treebank and explore
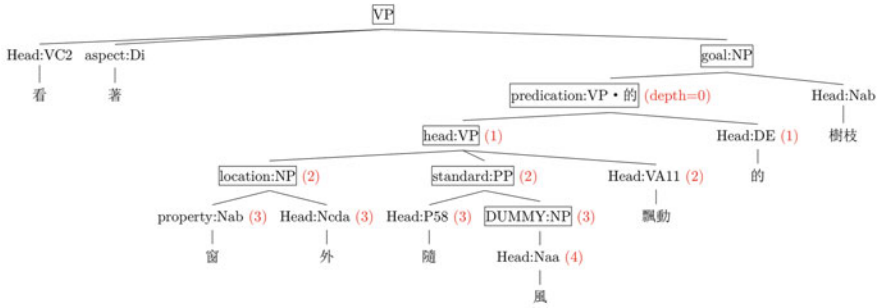
Fig. 23.1 Example of a Sinica tree structure of a relative clause, with depths in parentheses and phrasal nodes in boxes

the intricate connections between sentence comprehension, sentence production, and linguistic representation.

23.3 Distributional Regularities of Chinese Relative Clauses in the Sinica Treebank

Chinese relative clauses were extracted from the Sinica Treebank 3.0, which is based on the Sinica Corpus (http://asbc.iis.sinica.edu.tw/; Chen et al. 1996), a balanced corpus of contemporary Chinese texts produced between 1981 and 2007 (Huang et al. 2017). The Sinica Treebank 3.0 is composed of 361,834 words automatically parsed into 61,087 syntactic trees, which were manually checked and corrected before public release. Our corpus searches targeted NPs that contained prenominal modifier phrases headed by 的 *de* where the prenominal modifier contained a clause, a verb phrase (VP), or a verb. A sample tree diagram is provided in Fig. 23.1.

Our search yielded 3081 tokens, which were manually coded based on various syntactic and semantic properties of the head nouns, the prenominal clauses, and the location of complex NPs in the matrix clauses. The coding process was carried out and reviewed by native speakers of Standard Chinese (i.e., Mandarin), including both authors and several linguists. The coding guidelines were established by the first author. Cases where *de* served as a genitive marker (e.g., 人性的黑暗面 *renxing de heianmian* "the dark side of human nature") or appeared as part of an idiom (e.g., 所謂的 *suowei de* "so-called") as well as cases that contained incomplete RC fragments were excluded from further analysis ($N = 106$, 3% of all tokens). As a result, 2975 RCs were retained for subsequent analyses.

In addition to manually coding the syntactic and semantic properties of the RCs, we extracted the parts-of-speech (POS) tags of the embedded verbs based on verb classification in the Sinica Corpus and measured the syntactic complexity of the

embedded clauses based on several metrics.[1] These metrics included (a) the length of the prenominal RCs in terms of the number of characters and number of words, (b) the syntactic depth of the prenominal clauses in terms of the number of syntactic layers, and (c) syntactic complexity in terms of the number of phrasal nodes in the prenominal clauses. We will use Fig. 23.1 above to illustrate these measures.

The number of syllables or characters is the most straightforward measure. In Fig. 23.1, the prenominal clause contains seven characters/syllables, including the relativizer *de.* In Standard Chinese, the number of syllables/characters is almost equivalent to the number of morphemes. Phonological lengths thus quite closely reflect the amount of lexical content. The number of words (six in Fig. 23.1) is based on word segmentation in the Sinica Corpus. The number of layers (or depth) of a prenominal clause indicates how deep the clause is, which is measured by the number of edges on the path from the head (VP·的 in Fig. 23.1) to its deepest word (Head:Naa 風). Note that we counted from the head node of the RC (VP·的), not the head node of the whole tree fragment (VP), so in Fig. 23.1, the number of edges on the path is four. Tokens where more than one RC was found were excluded from this analysis. An additional measure of syntactic complexity is the number of phrasal nodes, whereby all non-terminal (non-leaf) nodes are counted. In the tree in Fig. 23.1, the embedded clause has four phrasal nodes—head:VP, location:NP, standard:PP, and DUMMY:NP. These phrasal nodes are roughly equivalent to the constituents in the sentence, which we believe are a good indicator of RC complexity.

The RCs were classified into six distinct types, with a primary focus on how the head nouns are reconstructed in the embedded clauses. Head nouns can be modified by clauses that are devoid of missing arguments. These RCs are gapless and are integrated with the head nouns as clausal complements (see Sect. 23.3.1). In most cases, the embedded clause contains a missing argument, with which the head noun is identified. A complete clause can be reconstructed by interpreting the missing argument as being coreferential with the head noun. In these instances, a filler-gap dependency exists between the head and the missing argument. We considered five subtypes where the head holds a dependency with an NP in the subordinate clause. In possessive RCs, the head is coreferential with the possessor argument of an embedded NP. In descriptive RCs, the head serves as the NP that the descriptive RC predicates on. The remaining three subtypes of RCs contain more obvious missing arguments in the embedded clause. In passive RCs, the head noun is coreferential with the missing subject NP of the embedded passive clause. In SRCs, the head noun is coreferential with the subject NP in the embedded clause. Finally, in ORCs, the head noun is coreferential with the object NP in the embedded clause. Table 23.1 provides definitions for the six types of RCs, each of which will be introduced in more detail. Furthermore, their respective distributions in the corpus will be discussed in subsequent sections:

---

[1]The python script is available at https://github.com/huhailinguist/processSinicaTree. Accessed on 13 September 2023

Table 23.1 Definitions of the relative clause types

RC type	Definition	Example
1. Gapless RC	Subordinate clauses that do not contain a missing argument of the embedded verb	七十萬人居住的以色列境內各阿拉伯城鎮 "the Arabic cities inside Israel where 700,000 people live"
2. Possessive RC	Subordinate clauses where a noun phrase serves as the possessee of the head noun. Usually, these possessee NPs form a part/whole or kinship relation with the head noun and subcategorizes for the head nouns as their inalienable possessor argument	一位身材魁梧、手持鐵椎的大力士 "a strong guy whose figure is stout and whose hand holds a hammer"
3. Descriptive RC	Subordinate modifiers that are headed by stative intransitive verbs, which can usually be modified by an intensifier like 很 *hen* "very"	年輕的一代 "the young generation"
4. Passive RC	Subordinate clauses that contain a passive structure headed by *bei* and a missing subject argument	被列為觀光區的原住民部落 "the aboriginal sites that have been designated as tourist districts"
4. SRC	Subordinate clauses where the subject argument of the embedded verb is empty and coreferential with the head noun	唱歌的小河 "the river that sings"
6. ORC	Subordinate clauses where the object argument of the embedded verb is empty and coreferential with the head noun	人類共同追求的目標 "the goal that all mankind pursues"

Figure 23.2 presents the percentile distributions of the different types of RCs. The majority (87%) of the RCs fell within two types of gapped RCs—SRCs (53%) and ORCs (34%), with SRCs outnumbering ORCs. The embedded clauses clearly showed the tendency of having missing subject or object arguments that were coreferential with the head nouns.

To get an initial glimpse of the complexity of the prenominal clauses, Table 23.2 shows the clausal lengths in terms of syllables/characters and words, the syntactic depths, and the syntactic complexity of the six types of RCs. The overall pattern was consistent across all four metrics (ps < 0.05, paired comparisons with Tukey correction). Descriptive RCs were the shortest and least complex, while passive RCs were the longest and most complex. SRCs were longer and more complex than ORCs.

Given that the syntactic category of the embedded verb plays an important role in selecting arguments, we further extracted the POS of the main verbs in the embedded clauses based on verb classification in the Sinica Corpus (Huang et al. 2017). The distribution of verb classes in the different RC types is presented in Table 23.3. The following sections will further discuss the POS properties of the different RC types using the information in Table 23.3.
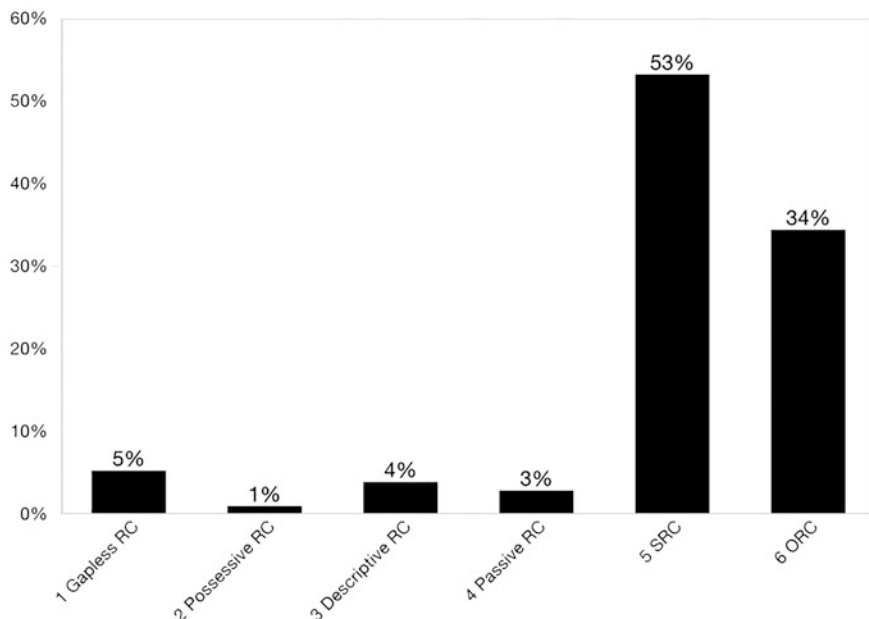
Fig. 23.2 Percentile distribution of relative clauses

Table 23.2 Length and complexity of relative clauses

	Gapless RC	Possessive RC	Descriptive RC	Passive RC	SRC	ORC
RC length: Number of characters	7.03	6.84	4.65	8.01	7.66	6.99
RC length: Number of words	4.37	4.48	3.09	5.30	4.70	4.40
Depth: Number of layers	3.10	3.09	2.23	3.85	3.24	3.04
Number of phrasal nodes	2.53	2.57	1.47	3.61	2.53	2.32

23.3.1 *Gapless Relative Clauses and Possessive Relative Clauses*

Both gapless RCs, exemplified in (5) to (7) below, and possessive RCs, as illustrated in (8), present themselves as complete clauses without obvious missing arguments or gaps. This section will distinguish these two types of RCs and compare their distributions in the corpus. Gapless RCs encompass three distinct types of compositional relations between the head noun and the embedded clauses. When the head noun functions as a relational noun (e.g., "time" and "space"), it takes an event argument and the prenominal clause fulfills the event argument requirement of the relational noun and serves as a clausal complement of the head noun. RCs like (5) are

Table 23.3 POS distributions of the embedded verbs of relative clauses (the most common categories are boldfaced)

POS	Gapless RC (%)	Possessive RC (%)	Descriptive RC (%)	Passive RC (%)	SRC (%)	ORC (%)
VA: Active intransitive verb	10	4	4	0	15	3
VB: Active pseudo-transitive verb	0	0	0	4	1	1
VC: Active transitive verb	**40**	26	9	**35**	**34**	**49**
VD: Ditransitive verb	4	0	0	0	1	4
VE: Active verb with a sentential object	6	0	0	6	4	11
VF: Active verb with a verbal object	2	0	0	5	1	1
VG: Classificatory verb	2	13	5	16	8	2
VH: Stative intransitive verb	17	**52**	**57**	11	7	7
VI: Stative pseudo-transitive verb	2	0	7	0	1	1
VJ: Stative transitive verb	10	4	9	15	17	9
VK: Stative verb with a sentential object	4	0	5	5	5	8
VL: Stative verb with a verbal object	2	0	5	1	6	2

commonly referred to as gapless relative clauses (Tsai 1997; Zhang 2008) or adjunct relative clauses (Lin 2018) in the literature. Gapless relative clauses also encompass *sloppy* relative clauses like (6), where the head noun is coerced into a relational noun, and it becomes integrated with a clausal complement to arrive at a sense of aboutness—akin to the function of "of" in English (Cheng and Sybesma 2005). Additionally, appositive relative clauses, exemplified by (7), fall under the category of gapless RCs. Together, gapless relative clauses accounted for approximately 5% of the relative clauses found in the Sinica Treebank.

(23.5) 七十萬人居住的以色列境內各阿拉伯城鎮

qishiwan__ren__juzhu__de__yiselie__jingnei__ge__alabo__chengzhen

700,000__people__live__DE__Israel__inside__each__Arabic__city

the Arabic cities inside Israel where 700,000 people live

(23.6) 昨日盤面拉高出貨的味道濃厚

zuori__panmian__lagao__chuhuo__de__weidao__nonghou

yesterday__stock.index__rise__sell__DE__taste__strong

The feel of stocks rising and being sold was strong yesterday.

(23.7) 民不與官鬥的道理

(continued)

min__bu__yu__guan__dou__de__daoli
civilian__not__with__ government.officials__fight__DE__principle
the principle that civilians should not fight against government officials

In contrast, in some gapless prenominal clauses, the head noun is non-relational and does not take the entire embedded clause as its complement or argument. Instead, the head noun forms a possessive association with a nominal argument located within the embedded clause. These RCs are classified as possessive RCs, as shown in (23.8) below. In these instances, the head noun is interpreted as the possessor argument of an embedded inalienable noun (e.g., *shencai* "figure" and *shou* "hand") (following Lin 2011). Possessive RCs constituted only 1% of the relative clauses extracted from the Sinica Treebank.

(23.8)   一位身材ᵢ魁梧、手ᵢ持鐵椎的大力士ᵢ
yi__wei__shencai$_i$__kuiwu__shou$_i$__chi__tiechui__de__dalishi$_i$
one__CL__figure$_i$__stout__hand$_i$__hold__hammer__DE__strong.guy$_i$
a strong guy whose figure is stout and whose hand holds a hammer

Distinctive reading patterns have been observed in gapless relative clauses like those in (23.5) to (23.7) and possessive relative clauses like (23.8) (Lin 2018) owing to the head nouns holding different dependency relations with the embedded clauses. Since the entire gapless RC is integrated with the adjunctive relational head noun, the complexity and frequency of the prenominal clause influence the processing difficulty of the complex NP. Conversely, the comprehension of possessive RCs is sensitive to the structural position of the dependent noun (possessee) in the prenominal clause. Dependent nouns located at subject positions as seen in (23.8) are generally easier to comprehend than those at lower syntactic positions such as objects. Gapless and possessive relative clauses are otherwise comparable in terms of pronominal clause lengths and syntactic complexity, and the lengths of the head nouns. All instances of possessive RCs found in our study involved an inalienable noun located in the subject position like in (23.8).

Furthermore, the animacies of the head nouns were distinctive between the two types of RCs. The majority (97%) of the head nouns in the gapless RCs were non-human relational nouns, while 53% of the head nouns in the possessive RCs were human possessors. Comparing the main verbs in gapless RCs and those in possessive RCs, it was observed that over half (52%) of the main verbs in the possessive RCs were stative intransitive verbs (VH), suggesting that possessive RCs mainly serve the function of describing the individual-level properties of the human head nouns.

23.3.2 Subject and Object Relative Clauses: Matrix Position, Animacy, and Complexity

The most common relative clauses are those where the head noun is interpreted as a key argument of the main verb in the embedded clause. These relative clauses typically contain a missing argument that is coreferential with the head noun. The highest grammatical functions in the Keenan-Comrie Accessibility Hierarchy (Keenan and Comrie 1977) shown in (23.9) below, namely, the subject and the object, are also the positions most frequently relativized in Chinese. These two types of relative clauses (not including descriptive SRCs and passive SRCs) account for over 87% of the relative clauses in the Sinica Treebank.

(23.9) Keenan-Comrie Accessibility Hierarchy (1977: 66):

subject > direct object > indirect object > oblique NP > genitive NP > object of comparison

---

Our study classified RCs that involved subject extraction into three subtypes: subject relative clauses that contain a missing subject argument (53%) like in (23.10) below, RCs that contain a passive structure (3%) like in (11), and prenominal modifiers that involve descriptive predicates (4%) like in (23.12). A typical RC that involved the extraction of a noun from an object position (34%) is exemplified by (23.13) below.

(23.10) [GAP$_i$]唱歌的小河$_i$

[GAP$_i$]__changge__de__xiaohe $_i$

[GAP$_i$]__sing__DE__river $_i$

the river that sings

(23.11) [GAP$_i$]被列為觀光區的原住民部落$_i$

[GAP$_i$]__bei__liewei__guangguangqu__de__yuanzhumin__buluo$_i$

[GAP$_i$]__BEI__designate.as__tourist.district__DE__aboriginal__site$_i$

the aboriginal sites that have been designated as tourist districts

(23.12) [GAP$_i$]年輕的一代$_i$

[GAP$_i$]__nianqing__de__yi__dai$_i$

[GAP$_i$]__young__DE__one__generation$_i$

the young generation

(23.13) 人類共同追求[GAP$_i$]的目標$_i$

renlei__gongtong__zhuiqiu__[GAP$_i$]__de__mubiao$_i$

mankind__together__pursue__[GAP$_i$]__DE__goal$_i$

the goal that all mankind pursues together

Passive relative clauses, with a word order like that in (23.14) below and an additional functional head such as 被 *bei*, 受 *shou*, 為 *wei*, 由 *you*, 遭 *zao*, etc., are distinctive from SRCs and ORCs. Notably, in the so-called "short passives", the agent NP may be absent, and the head noun typically assumes the role of the theme

Table 23.4 Distribution of relative clauses as a function of extraction types and position in matrix clauses

	S-SRC (%)	S-ORC (%)	O-SRC (%)	O-ORC (%)	Total
Wu et al. (2011)	39.5	26.5	21.3	12.7	347
Our study	31.8	19.2	26.2	22.8	1542

S-SRC SRC in matrix subject position, *S-ORC* ORC in matrix subject position, *O-SRC* SRC in matrix object position, *O-ORC* ORC in matrix object position

or patient NP of the embedded verb. Due to these distinctions, we have categorized passive RCs separately and will discuss their distributional properties in Sect. 23.3.3.

(23.14) [GAP$_i$]__*bei/zao/shou*__(Agent.NP)__Verb__DE__Patient.NP$_i$

Given that stative verbs in Chinese are typically predicative of subject NPs, as in (15) below, they can be regarded as RCs that involve subject extractions. However, they also diverge quite significantly from the typical gapped relatives like SRCs and ORCs, which entail the relativization of a key argument of the embedded verb. Based on the information provided in Table 23.2, descriptive relative clauses were notably shorter in length (averaging 4.65 characters) and displayed a higher degree of simplicity (averaging 1.47 phrasal nodes) compared to RCs that involved extractions from subject or object positions. They can thus be taken as simple predicates that are integrated with the head nouns without having to involve a structure-based filler-gap dependency, much like gapless relative clauses and adjectives in English. Notably, the embedded verbs in these descriptive RCs were mainly stative intransitive verbs (57% being VH verbs).

(23.15)   這些孩子還很年輕
zhe__xie__haizi__hai__hen__nianqing
this__CL__kids__still__very__young
These kids are still young.

We will now turn to the distributional properties of RCs that involve the extraction of subject and object arguments. As introduced, SRCs and ORCs are among the most commonly studied sentence structures. Of the relative clauses extracted from the Sinica Treebank, SRCs (53%) appeared more frequently than ORCs (34%), which is consistent with findings in other languages and in other studies on the Chinese language. Table 23.2 also shows that SRCs were longer and more complex than ORCs. Sentence comprehension studies on Chinese RCs have yielded a mix of SRC advantages and ORC advantages, as reviewed in Sect. 23.2. The corpus distributions suggest that Chinese language users may, on the whole, be more experienced with SRCs than ORCs.

One important discourse function of RCs is to reference information already present in the background and present the focused NP for predication. The RC's position in the matrix clause therefore plays a pivotal role for understanding the discourse functions. Typically, the subject position of a sentence imparts grounding

Table 23.5 Distribution of relative clauses as a function of extraction types, position in matrix clauses, and existence of *shi* in matrix predicates

	S-SRC (%)	S-ORC (%)	O-SRC (%)	O-ORC (%)	Total
N-*shi*-N	12.2	29.3	33.0	25.5	482
SV(O)	40.8	14.6	23.1	21.5	1060

S-SRC SRC in matrix subject position, *S-ORC* ORC in matrix subject position, *O-SRC* SRC in matrix object position, *O-ORC* ORC in matrix object position

information shared by interlocutors whereas the object position provides new and focused information. Sentence processing research has revealed that, overall, Chinese RCs are more frequently expected in the subject position (Lin 2012). Table 23.4 summarizes the findings of Wu et al. (2011), who extracted 1218 relative clauses from the first 1000 files in the Chinese Treebank 5.0 (Xue et al. 2005), and compares them with the distributions in our study based on the Sinica Treebank.

The general distributions were similar in both studies, with RCs appearing more often in the subject positions of matrix clauses than in the object positions. Furthermore, there were more SRCs than ORCs in both positions. However, our study differs from Wu et al. (2011) in that the SRCs in our study were more inclined to modify matrix subject NPs, while the ORCs tended to modify matrix object NPs. This contrast was even more pronounced when we differentiated between matrix clauses that contained the presentative copula *shi* and those that did not, as shown in Table 23.5.

Sentences containing non-*shi* predicates presented a stronger tendency for an SRC to modify a subject NP (41% vs. 23%) and for an ORC to modify an object NP (22% vs. 15%). This interplay between the presence of the presentative copula *shi* and the distribution of RCs in matrix clauses underscores the importance of differentiating sentences containing *shi* and those that do not when studying RC positions. It also implies that the grammatical function of the head noun in the RC interacts with its function in the matrix clause. When considering sentences without *shi* it becomes apparent that head nouns tend to fulfill the same grammatical functions in both the subordinate and matrix clauses. This observation can be explained by two plausible accounts. First, in terms of production, it may be more efficient to maintain consistent grammatical functions in both the embedded clause and the matrix clause. Secondly, this distribution also aligns with the general semantic tendency that NPs in the subject position tend to be human and those in the object position tend to be non-human entities, as proposed by Traxler et al. (2002). The humanness/animacy factor can lead to the tendency for the heads of SRCs to be human nouns, which are also preferably located in the subject position of the matrix clause. On the other hand, the heads of ORCs are more likely to be inanimate and preferably located in the object position of the matrix clause.

To delve deeper into these two accounts, we further conducted an analysis of the animacy distribution of the head nouns in relation to the types of grammatical extractions (SRC vs. ORC) and their matrix positions (Subject vs. Object). In terms of animacy and humanness, the head nouns were classified into five categories,

Table 23.6 Examples and distribution of head noun animacy/humanness

Animacy/humanness	Example	%
Inanimate	所費的功夫 "the effort it takes"	58.0
Human	她教過的學生 "the students that she taught"	37.0
Animal	觀賞到的動物 "the animals that people see"	3.1
Plant	這種種子所長成的草 "the kind of grass that this kind of seed turns into"	1.1
Metaphorical animate	唱歌的小河 "the stream that sings"	0.5
Unclassifiable or mixed		0.3
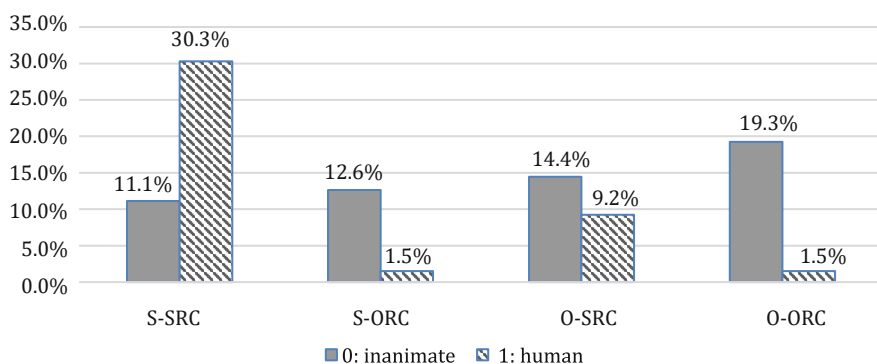
Fig. 23.3 Percentile distribution of head noun animacy/humanness, RC matrix positions, and RC types

as shown in Table 23.6. We focused on the distribution of inanimate NPs (58%) and human NPs (37%) because these two categories accounted for the majority (95%) of the data.

Figure 23.3 presents the percentile distribution of SRCs and ORCs ($N = 997$, where the head noun is either inanimate or human) as a function of head noun animacy/humanness and matrix positions, excluding the matrix sentences that contained *shi*.

The distribution percentages shown in Fig. 23.3 affirm the overall animacy/ humanness asymmetry in terms of grammatical positions, which has been observed across languages (Fox and Thompson 1990). Specifically, subject positions are more likely to be occupied by human nouns and object positions are more likely to be occupied by inanimate nouns. This asymmetry was also evident in RC extraction types, as both SRCs and ORCs showed distinctive animacy preferences.

As shown in Fig. 23.3, in the matrix subject position, while SRCs mainly modified human NPs, only very few ORCs modified human NPs. In the matrix object position, the proportion of inanimate NPs increased for both SRCs and ORCs and the proportion of human NPs decreased, especially in SRCs. The animacy
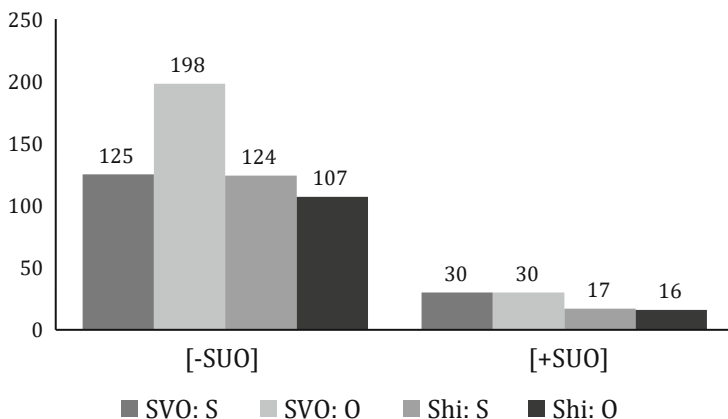
Fig. 23.4 Distribution of RC matrix positions as a function of the existence of *suo* in ORCs (numbers indicate instances)

preference within the matrix clauses and that within the embedded clauses presented an intriguing interaction, resulting in a competition between the two levels of grammatical functions based on their animacy preferences. In the matrix subject position, the animacy preference of the embedded RC type determined the tendency, while in the matrix object position, that of the matrix clause determined the tendency. In both matrix positions, ORCs modified inanimate head NPs more frequently than SRCs, while SRCs featured a higher proportion of human head nouns than ORCs only in the matrix subject position.

Regarding the POS of the embedded verbs (see Table 23.3), in both SRCs and ORCs, transitive action verbs (VC) were the most common. The different verb classes were fairly evenly distributed in SRCs but more skewed toward transitive verbs that required an object argument in ORCs. Compared with ORCs, SRCs had more intransitive action verbs like *pao* "to run" (VA: 15%) that required only one subject argument, classification verbs like *xing* "to be named as" (VG: 8%), and stative verbs that required only one object argument like *daibiao* "to stand for" (VJ: 17%).

Finally, ORCs in Standard Chinese are known to sometimes appear with the particle *suo* located before the main verb, as in (16) below, which is associated with greater formality and literary style. Among the RCs in our study, 164 RCs (5.5%) featured the particle *suo*. ORCs with *suo* were longer than those without *suo* in the embedded clauses (9.7 vs. 6.8 characters, $t = 5.92$, $p < 0.001$), which is consistent with the notion that constituent length serves as an indicator of formality in Standard Chinese, with longer constituents generally conveying a higher degree of formality.

(23.16)	專家所具備的投資能力比一般人高。
zhuanjia__suo__jubei__de__touzi__nengli__bi__yiban__ren__gao
expert__SUO__have__DE__invest__ability__compare__regular__person__high
The ability to invest that experts have is higher than that of regular people.

Upon comparing ORCs with *suo* and ORCs without *suo* in terms of whether they modify a matrix subject NP or a matrix object NP in Fig. 23.4, a noteworthy observation emerged. While ORCs without *suo* tended to appear in the matrix object position in sentences that did not involve *shi*, ORCs with *suo* are equally distributed in subject and object positions. This finding suggests that the enhanced formality associated with *suo* in an ORC overrides the animacy propensity that was discussed above and leads to a more balanced appearance of an ORC in the subject and object positions of matrix clauses.

23.3.3   Passive Relative Clauses

Due to the increased syntactic complexity associated with an additional functional head (e.g., *bei*), passive RCs are longer and more complex than SRCs and ORCs (see Table 23.1). Passive RCs, as exemplified in (11), stand between SRCs and ORCs as a third category that involves the relativization of a key argument associated with the embedded verb. In terms of thematic content, passive RCs are similar to ORCs as it is the patient NP of the embedded clause that is relativized. In terms of the grammatical position of the relativized gap, a passive RC is more similar to an SRC, where the gap is located in the subject position.

We looked at the position of passive RCs in matrix clauses and found that, like ORCs, the majority (69%) of passive RCs were located at the matrix object position. Further exploring the animacy distribution of the head nouns in SRCs, ORCs, and passive RCs, as shown in Table 23.7 below, based on the coding scheme in Table 23.6, we found that passive RCs were more similar to ORCs, with the head noun more likely to be an inanimate NP, though the tendency of having an inanimate head noun was not as strong as that of ORCs. These observations suggest that relativized patient NPs tend to be inanimate nouns. Moreover, ORCs and passive RCs were similar in terms of their thematic content and animacy preferences.

On the other hand, there appeared to be more human head nouns in passive RCs (31%) than in ORCs (7%), suggesting that a human patient noun is more likely to be relativized if it appears in the subject position of a passive clause than if it appears in the object position of an SVO clause. This finding suggested that passivization promoted the saliency of a patient NP for relativization.

Table 23.7   Animacy distribution of SRCs, ORCs, and passive RCs

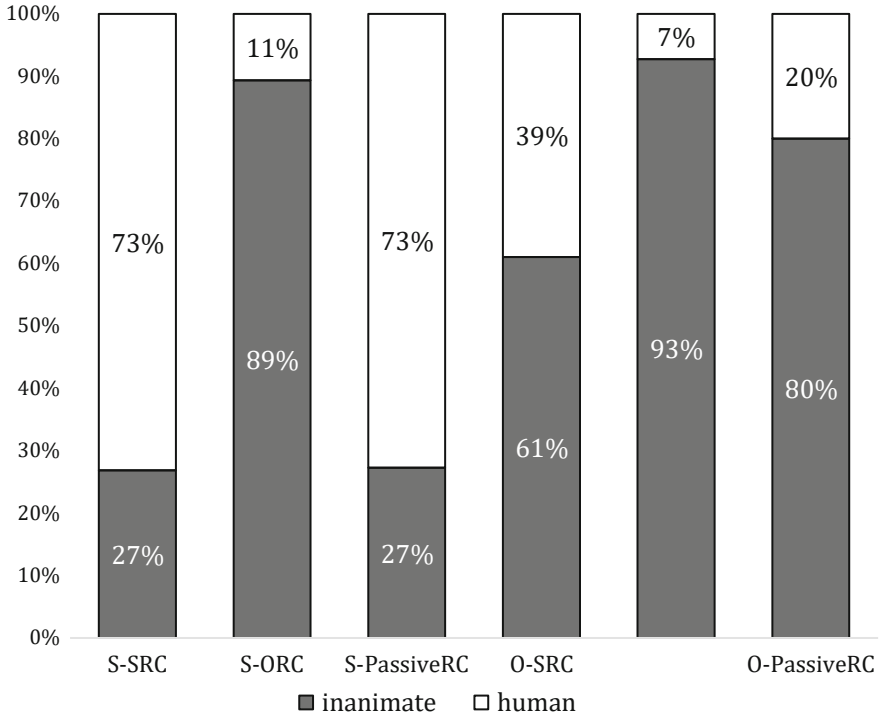| Animacy of head	SRC (%)	ORC (%)	Passive RC (%)
0: Inanimate	40	89	58
1: Human	55	7	31
2: Animal	4	2	8
3: Plant	1	2	3
4: Metaphorical animate	1	0	0
5: Unclassifiable or mixed	0	1	1

Fig. 23.5 Percentile distribution of animacy/humanness, RC matrix positions, and RC types

Turning now to the interplay between the animacy of the head noun and the position of the complex NP in the matrix clause in Fig. 23.5, the head nouns of passive RCs were predominantly human NPs in matrix subject positions (i.e., S-PassiveRC) but inanimate NPs in matrix object positions (i.e., O-PassiveRC). This distribution again confirms that passive RCs fall between SRCs and ORCs. The animacy of its head noun mirrors that of an SRC in the matrix subject position but aligns more closely with that of an ORC in the matrix object position.

The POS distribution of the embedded verbs in passive RCs was similar to those of ORCs. Unlike SRCs, passive RCs did not contain any intransitive action verbs (VA) and had more instances of classification verbs like *chengwei* "to call" (VG) serving as the main verb. In terms of thematic ordering, passive RCs presented the canonical order of Agent-Verb-Patient, similar to that of an ORC. Lin (2015) compared the reading patterns of passive RCs, RCs that involved the disposal marker *ba*, as shown in (23.17) below, and normal SRCs. The study's findings indicated that passive RCs exhibited the shortest reading times. This outcome underscores the importance of thematic ordering in processing relative clauses.

(23.17) [GAP$_i$]__*ba*__Patient.NP__Verb__DE__Agent.NP$_i$

23.4 Classifier Position in Relative Clauses

One important function of RCs in discourse is to serve the restrictive function; that is, RCs help bring attention to particular referents already present in the background knowledge. One well-known proposal about how restrictiveness is expressed in Standard Chinese focuses on the position of the determiner-classifier phrase in relation to the relative clause (Chao 1968). When a relative clause precedes a determiner-classifier phrase, as in (18a) below, it is considered restrictive because the pre-determiner-classifier position is an edge-position that marks focus. When a relative clause appears after a determiner-classifier phrase, as in (18b), it lacks the focus marking and can be interpreted either as restrictive or non-restrictive (Lin 2012).

(23.18a) 他在台北拇指山下許的那個願
ta__zai__taibei__muzhishan__xuxia__de__na__ge__yuan
he__at__Taipei__Mt.Muzhi__make__DE__that__CL__wish
the wish that he made on Mt. Muzhi in Taipei
(23.18b) 這場可能贏的球
zhe__chang__keneng__ying__de__qiu
this__CL__likely__win__DE__ball.game
the ball game that (I am) likely to win

Over the years, this proposal has sparked controversy. One way to test Chao's (1968) proposal is to examine whether the position of determiner-classifier (CL) phrases interacts with the matrix positions of complex NPs since restrictive relative clauses are more likely to appear in subject positions to ground referents (Gibson et al. 2005). Following this logic, we expected to find more occurrences of RCs that appeared before classifier phrases than RCs that appeared after classifier phrases in subject positions.

Focusing on the matrix positions of RCs that co-occurred with classifier phrases ($N = 174$) and distinguishing sentences that contained *shi* from those that did not, we found that RCs were generally more likely to appear after CL phrases, except when they appeared in the subject position of a sentence containing *shi* (see Fig. 23.6).

The finding that RCs were, overall, more likely to appear after classifiers suggests that the post-classifier position (i.e., CL-RC) is an unmarked position for RCs. The greater occurrences of RCs in the pre-classifier position when complex NPs appeared in the matrix subject position of a sentence with *shi* suggest that (i) the subjects in sentences with *shi* are preferred for grounding referents and (ii) RCs appearing in pre-classifier positions are indeed more likely to be used in a restrictive sense. These findings are consistent with Chao's (1968) proposal and further specify that grounding most likely happens in the subject position of a sentence containing *shi*.
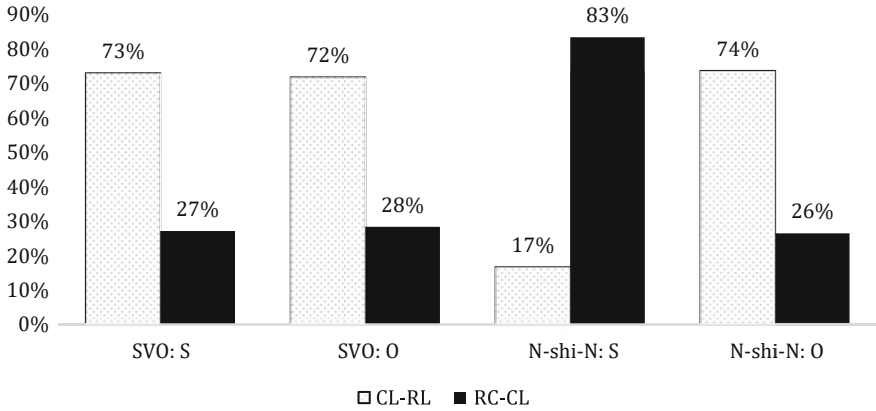
Fig. 23.6 Position of RCs in relation to CL phrases as a function of matrix positions
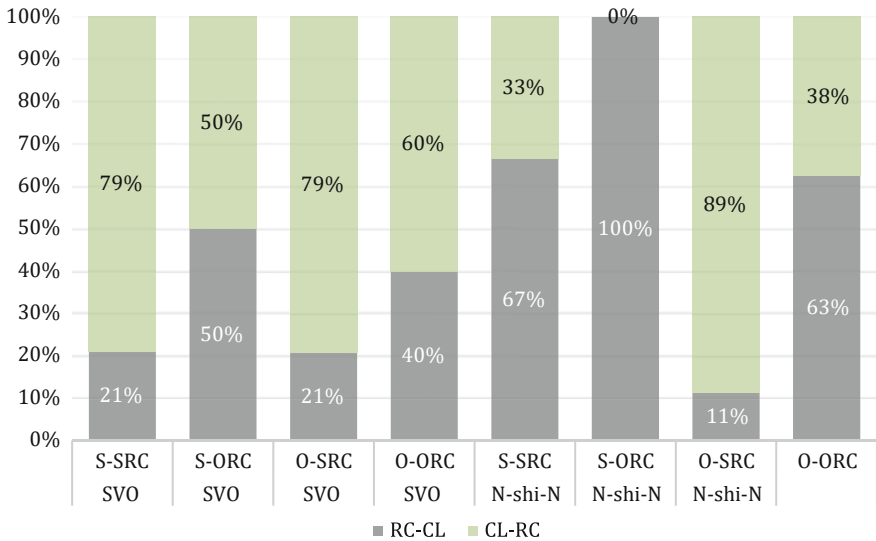

Fig. 23.7 Position of RCs and CL phrases as a function of matrix positions and RC types

Figure 23.7 further breaks down the distributions in Fig. 23.6 as a function of RC types (SRCs vs. ORCs) and shows an overall trend where ORCs appeared in the marked pre-classifier position more often than SRCs did.

To understand this finding, we schematically sketched the linear sequencing of classifier phrases in relation to RCs in (19) below:

(23.19a)	CL-SRC:	CL [_ V N1 de] N2
(23.19b)	CL-ORC:	CL [N1 V _ de] N2
(23.19c)	SRC-CL:	[_ V N1 de] CL N2
(23.19d)	ORC-CL:	[N1 V _ de] CL N2

Our observation in Fig. 23.7 was that, relative to (19b) and (19a), respectively, (19d) appeared more often than (19c), which suggests the possibility that language users may have attempted to avoid the potential classifier-noun clash in the CL-ORC condition in (19b) by moving the ORC to a pre-classifier position—assuming that a decision was made between (19a) and (19c) as well as between (19b) and (19d). Wu (2011) similarly found less than 5% of classifier phrases before ORCs in the corpus and interpreted this as a production strategy to avoid ambiguity between the classifier and the first noun in the relative clause. Interestingly, passive RCs where no semantic clash exists after the classifier displayed the same pattern as SRCs in preferring the unmarked position (88% vs. 12%) in classifier phrases, further supporting the notion of ambiguity avoidance in ORCs.

Being prenominal, Chinese RCs often present a challenge for comprehension because they can initially be taken as a matrix clause (Lin and Bever 2011). Sentence comprehension research has used pre-RC classifiers as a cue for marking constituent boundaries. In (20) below, because the classifier 塊 *kuai* and the following pronominal 他 *ta* "he" cannot form a local constituent, a phrasal boundary must be created between the two, signaling *ta* as the beginning of the embedded clause. This boundary has been employed as a cue that may indicate the beginning of an embedded clause for sentence comprehension (e.g., Lin 2018). Based on the production data from the corpus, however, ORCs rarely appeared after determiner-classifier phrases.

(23.20)	一塊他喜歡[GAP$_i$]的石頭$_i$
	yi__kuai__ta__xihuan__[GAP$_i$]__de__shitou$_i$
	one__CL__he__like__[GAP$_i$]__de__rock$_i$
	a rock that he likes

23.5 Headless Relative Clauses

The head nouns of RCs can be left empty, as in (21) below, when they can be easily reconstructed from context or are of generic nature. Among the collected tokens, 303 RCs (10%) were headless. The majority (95%) of headless RCs were either SRCs or ORCs. Interestingly, in contrast to the overall distribution of RC types where we found more SRCs than ORCs (see Fig. 23.1), headless RCs were more often found in ORCs (58%) than in SRCs (37%) (see Table 23.8). This pattern suggests that head nouns that are coreferential with the object of the embedded

Table 23.8 Distribution of headless RCs as a function of RC types

Gapless RC	Possessive RC	Descriptive RC	Passive RC	SRC	ORC
0.99%	2.31%	0.66%	0.99%	37.29%	57.76%
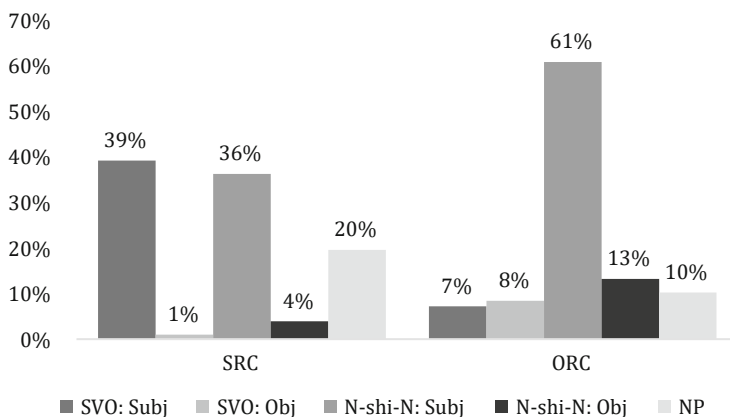
Fig. 23.8 Position of headless RCs as a function of matrix positions

clause are more likely to be omitted given their lower saliency in discourse. These omitted head nouns were more likely to be inanimate (65%) than human (30%).

(23.21)	所以真正賺錢的都是這些廠商
	suoyi__zhenzheng__zhuanqian__de__dou__shi__zhe__xie__changshanng
	therefore__really__make.profit__DE__DOU__SHI__this__CL__merchant
	Therefore, those who really make a profit are the merchants.

Focusing on SRCs and ORCs, headless SRCs (20%) were more likely to appear as an independent topicalized NP than headless ORCs (10%). In matrix clauses, headless SRCs appeared more often in the subject position, which is consistent with the overall preference for SRCs to appear in the matrix subject position (see Fig. 23.8). Interestingly, headless ORCs did not show the same preference for the matrix object position. For sentences with *shi*, in particular, headless ORCs were more likely to appear in the subject position. The tendency for a headless RC to appear in the subject position of a sentence with *shi* suggests that headless RCs are mainly used for grounding referents that already exist in the background.

23.6 Concluding Remarks

This chapter presented topics on the comprehension of Chinese relative clauses in relation to the distributional properties of Chinese relative clauses in the Sinica Treebank. The data were analyzed regarding structural dimensions such as the length and complexity of relative clauses, their positions in matrix clauses, semantic dimensions regarding the animacy of head nouns, and the position of classifier phrases in relation to relative clauses. These corpus data, which serve as a snapshot of collective sentence production, have contributed to our understanding of the relation between production and comprehension. Moreover, they have raised intriguing questions about sentence processing for further exploration.

References

Bever, Thomas G. 1970. The cognitive basis for linguistic structures. *Cognition and the Development of Language* 279(362):1–61.

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. University of California Press.

Chen, Keh-Jiann, Chu-Ren Huang, Li-Ping Chang, and Hui-Li Hsu. 1996. Sinica corpus: Design methodology for balanced corpora. In *Proceedings of the 11ᵗʰ Pacific Asia Conference on Language, Information and Computation*, ed. Byung-Soo Park and Jong-Bok Kim, 167–176. Seoul, Korea.

Chen, Keh-Jiann, Chu-Ren Huang, Feng-Yi Chen, Chi-Ching Luo, Ming-Chung Chang, Chao-Jan Chen, and Zhao-Ming Gao. 2003. Sinica treebank: Design criteria, representational issues and implementation. In *Building and using parsed corpora*, ed. Anne Abeille, 231–248. Dordrecht: Kluwer.

Chen, Zhong, Lena Jäger, and Shravan Vasishth. 2012. How structure-sensitive is the parser? Evidence from Mandarin Chinese. In *Empirical approaches to linguistic theory: Studies of meaning and structure*, 43–62. Berlin: Mouton de Gruyter.

Cheng, Lisa Lai-Shen, and Rint Sybesma. 2005. A Chinese relative. In *Organizing grammar: Linguistic studies in honor of Henk van Riemsdijk*, ed. Hans Broekhuis, Norbert Corver, Rint Huybregts, Ursula Kleinhenz, and Jan Koster, 69–76. Berlin: Mouton de Gruyter.

Denes, Peter B., and Elliot Pinson. 1993. *The Speech Chain*. Macmillan.

Ferreira, Fernanda. 1991. Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language* 30(2):210–233.

Fox, Barbara A., and Sandra A. Thompson. 1990. A discourse explanation of the grammar of relative clauses in English conversation. *Language* 66:297–316.

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68(1): 1–76.

Gibson, Edward, and H.-H. Iris Wu. 2013. Processing Chinese relative clauses in context. *Language and Cognitive Processes* 28(1–2):125–155.

Gibson, Edward, and Tessa Warren. 2004. Reading time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax* 7(1):55–78.

Gibson, Edward, Timothy Desmet, Daniel Grodner, Duane Watson, and Kara Ko. 2005. Reading relative clauses in English. *Cognitive Linguistics* 16(2):313–353.

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*, 1–8. Pittsburgh, Pennsylvania.

Hale, John. 2006. Uncertainty about the rest of the sentence. *Cognitive Science* 30(4):643–672.

Hsiao, Franny, and Edward Gibson. 2003. Processing relative clauses in Chinese. *Cognition* 90(1): 3–27.

Huang, Chu-Ren, Shu-Kai Hsieh, and Keh-Jiann Chen. 2017. *Mandarin Chinese words and parts of speech: A corpus-based study*. Taylor & Francis.

Jäger, Lena, Zhong Chen, Qiang Li, Chien-Jer Charles Lin, and Shravan Vasishth. 2015. The subject relative advantage in Chinese: Evidence for expectation-based processing. *Journal of Memory and Language* 79:97–120.

Keenan, Edward L., and Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8(1):63–99.

King, Jonathan, and Marcel Adam Just. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language* 30(5):580–602.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3):1126–1177.

Lin, Chien-Jer Charles. 2011. Processing (in)alienable possessions at the syntax-semantics interface. In *Interfaces in Linguistics: New research perspectives*, ed. Raffaella Folli and Christiane Ulbrich, 351–367. New York: Oxford University Press.

Lin, Chien-Jer Charles. 2012. Restrictiveness and information status of Chinese relative clauses: Evidence from discourse comprehension. Paper presented at the *Pragmatics Festival*. Bloomington, Indiana.

Lin, Chien-Jer Charles. 2013. Effects of syntactic complexity and animacy on the initiation times for head-final relative clauses. Poster presented at the *26th Annual CUNY Conference on Human Sentence Processing*. Columbia, South Carolina.

Lin, Chien-Jer Charles. 2014. Effect of thematic order on the comprehension of Chinese relative clauses. *Lingua* 140:180–206.

Lin, Chien-Jer Charles. 2015. Thematic orders and the comprehension of subject-extracted relative clauses in Mandarin Chinese. *Frontiers in Psychology* 6:1255. (*Special research topic on encoding and navigating linguistic representations in memory*, ed. Claudia Felser, Colin Phillips, and Matthew Wagers).

Lin, Chien-Jer Charles. 2018. Subject prominence and processing filler-gap dependencies in prenominal relative clauses: The comprehension of possessive relative clauses and adjunct relative clauses in Mandarin Chinese. *Language* 94:758–797.

Lin, Chien-Jer Charles, and Thomas G. Bever. 2006. Subject preference in the processing of relative clauses in Chinese. In *Proceedings of the 25th West Coast Conference on Formal Linguistics*, 254–260. Somerville, Massachusetts.

Lin, Chien-Jer Charles, and Thomas G. Bever. 2011. Garden path in the processing of head-final relative clauses. In *Processing and producing head-final structures*, ed. Yuki Hirose, Hiroko Yamashita, Jerome Packard, 277–297.

Lin, Yow-Yu, and Susan Garnsey. 2011. Verb bias in Mandarin relative clause processing. *Concentric: Studies in Linguistics* 37(1):73–91.

MacDonald, Maryellen C. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology* 4:226.

O'Grady, William. 2011. Relative clauses: Processing and acquisition. In *The acquisition of relative clauses: Processing, typology and function*, ed. Evan Kidd, 13–38. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Packard, Jerome L., Zheng Ye, and Xiaolin Zhou. 2011. Filler-gap processing in Mandarin relative clauses: Evidence from event-related potentials. In *Processing and producing head-final structures*, ed. Yuki Hirose, Hiroko Yamashita, Jerome Packard, 219–240.

Qiao, Xiaomei, Liyao Shen, and Kenneth I. Forster. 2012. Relative clause processing in Mandarin: Evidence from the Maze Task. *Language and Cognitive Processes* 27:611–630.

Reali, Florencia, and Morten H. Christiansen. 2007. Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language* 57(1):1–23.

Roland, Douglas, Frederic Dick, and Jeffrey L. Elman. 2007. Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language* 57(3):348–379.

Sung, Yao-Ting, Jih-Ho Cha, Jung-Yueh Tu, Ming-Da Wu, and Wei-Chun Lin. 2016. Investigating the processing of relative clauses in Mandarin Chinese: Evidence from eye-movement data. *Journal of Psycholinguistic Research* 45:1089–1113.

Traxler, Matthew J., Robin K. Morris, and Rachel E. Seely. 2002. Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language* 47(1):69–90.

Tsai, Wei-Tien Dylan. 1997. On the absence of island effects. *Tsing Hua Journal of Chinese Studies* 27:125–149.

Wu, Fuyun. 2011. Frequency issues of classifier configurations for processing Mandarin object-extracted relative clauses: A corpus study. *Corpus Linguistics and Linguistic Theory* 7:203–227.

Wu, Fuyun, Elsi Kaiser, and Elaine Andersen. 2011. Subject preference, head animacy and lexical cues: A corpus study of relative clauses in Chinese. In *Processing and producing head-final structures*, ed. Yuki Hirose, Hiroko Yamashita, Jerome Packard, 173–94.

Xue, Naiwen, Fei Xia, Fu-Dong Chiou, and Marta Palmer. 2005. The Penn Chinese treebank: Phrase structure annotation of a large corpus. *Natural Language Engineering* 11(2):207–238.

Zhang, Niina. 2008. Gapless relative clauses as clausal licensers of relational nouns. *Language and Linguistics* 9:1005–1028.