When missing NPs make double center-embedding sentences acceptable

Nick Huang & Colin Phillips
znhuang@nus.edu.sg, colin@umd.edu

Nov 2020

(Word count: 16,540 words)

Abstract

A number of languages, such as English, exhibit a grammaticality illusion in ungrammatical double center-embedding sentences where a VP is missing. This article shows that the illusion generalizes to ungrammatical Mandarin Chinese double center-embedding sentences where the head NP of a relative clause is missing.

The Mandarin illusion raises interesting questions for existing accounts of center-embedding illusions. Mandarin missing NP sentences consist of three transitive verbs and only three NPs; the clear shortage of NPs should affect the thematic relations built for such sentences, with potential consequences for acceptability. We explore these issues with acceptability judgment experiments. We show that these illusory sentences receive distinct thematic interpretations compared to their better-studied missing VP counterparts, in ways not predicted by structural forgetting or interference accounts. A computational simulation further shows that the Mandarin illusion is problematic for accounts that attribute cross-linguistic variation in the illusion to differences in language experience.

To capture cross-linguistic variation, we build on existing interference accounts, in which the parser mis-attaches a verb or NP to the main clause instead of a relative clause. We supplement this approach with a repair process, in which the parser tracks thematic relations, repairing them where necessary so no verb or noun is "orphaned." We suggest that the illusion of grammaticality arises when the parser can establish thematic relations between all verbs and nouns. This interference-and-repair approach provides a unified analysis of the missing VP and missing NP illusions, while accounting for the observed difference in thematic relations.

Keywords

double center-embedding, sentence processing, grammaticality illusion, Mandarin Chinese

1 Introduction

Native speakers are usually very sensitive to grammatical violations. They can quickly tell when a sentence is ill-formed, even though they might not be able to give a precise description of the problem. However, for a subset of ungrammatical sentences, native speakers do not show the same kind of sensitivity. Because these "illusions of grammaticality" occur systematically, they are of psycholinguistic interest, as they reflect some systematic error in the way linguistic representations are built or accessed.

This article is concerned with a particularly acute type of grammaticality illusion associated with sentences with double center-embedding (DCE). In English, these sentences are formed by embedding an object relative clause inside another object relative clause that modifies a subject (1), producing a sequence of three noun phrases (NPs) followed by three verb phrases (VPs). While syntactically well-formed, DCE sentences are generally hard to parse and are judged as strongly unacceptable (Chomsky & Miller 1963; Miller & Isard 1964; Cowper 1976; Gibson & Thomas 1999; among others). Curiously, these sentences are perceived to improve in acceptability if a VP – specifically, the middle VP – is omitted (Frazier 1985, attributing the original observation to Janet Fodor; Gibson & Thomas 1999; see also Pulman 1986).

(1) Double center embedding (DCE) (Gibson & Thomas 1999)
[$_{NP}$ The novel] that [$_{NP}$ the horror author] who [$_{NP}$ the publishing company] [$_{VP}$ had recently fired] [$_{VP}$ had typed quickly] [$_{VP}$ was banned by the local library].

(2) Missing VP sentence
[$_{NP}$ The novel] that [$_{NP}$ the horror author] who [$_{NP}$ the publishing company] [$_{VP}$ had recently fired] [$_{VP}$ was banned by the local library].[1]

An important finding in recent work is that this effect, often referred to as a "missing VP illusion," varies across languages: sentences that are string-equivalent to (2) elicit the illusion in English and French but not in German and Dutch (Gimenes et al. 2009; Vasishth et al. 2010; Frank et al. 2016; Frank & Ernst 2019; see Häussler & Bader 2015 and Bader 2016 for discussion of exceptions in German). A number of recent proposals attribute this variation to cross-linguistic differences in the frequency of verb-final clauses and verb clusters, suggesting that linguistic experience can influence the processing of DCE sentences and the size of the illusion (Vasishth et al. 2010; Frank et al. 2016; Futrell & Levy 2017; Frank & Ernst 2019; Futrell et al. 2020).

This article is interested in why ungrammatical DCE sentences in some languages can exceptionally elicit a sense of acceptability and how speakers interpret these sentences. We bring in novel data from Mandarin Chinese. We report two acceptability judgment experiments, one on Mandarin DCE sentences and one on their English counterparts, replicating Gibson & Thomas 1999. We demonstrate a missing NP illusion for Mandarin that is analogous to the better-studied missing VP illusion.

The Mandarin illusion data also let us evaluate experience-based accounts about cross-linguistic variation. We adapt an existing computational model (Futrell et al. 2020, also Futrell & Levy 2017), extending it with the goal of modelling cross-linguistic variation in the illusion.

[1] For ease of reference, we sometimes label constituents by their linear position in the sentence string, following Gibson & Thomas (1999), as opposed to their hierarchical position. Hence, in English, constituents would appear in the following order: NP1-NP2-NP3-VP1-VP2(-VP3).

Our results pose challenges to existing accounts. Mandarin missing NP sentences are interpreted differently from their English missing VP counterparts, in ways not predicted by these accounts. Our modeling results suggest that experience-based accounts incorrectly predict that Mandarin should lack the illusion, like German and Dutch. We discuss how an interference-based account, such as Bader 2016, might be modified to capture cross-linguistic differences more easily.

This paper is structured as follows. In Section 2, we review the missing VP illusion and a number of existing accounts of this illusion. We then review center-embedding in Mandarin in Section 3 and present the experiments in Sections 4 through 7. We discuss implications of our results in Section 8, before concluding in Section 9.

2 The missing VP illusion

In this section, we consider why missing VP sentences in languages like English exhibit a grammaticality illusion. For scope reasons, we will review only recent memory-based accounts, which have been adapted to explain cross-linguistic variation.[2] We distinguish between three classes of accounts: structural forgetting, language experience, and interference accounts.

Existing accounts are mostly focused on why a VP is omitted in DCE sentences in some, but not all, languages. We can break down this question into two smaller questions: (i) What is the representation(s) in which a VP is omitted, and over which acceptability is computed? (ii) How does a VP get omitted in that representation in some (but not all) languages? The first question is often addressed obliquely, so where necessary, we spell out what the underlying assumptions might be. We also discuss how these accounts might handle cross-linguistic variation, although not all of them were explicitly developed for this issue.

After reviewing these two questions, we discuss the perception of acceptability: what aspect of the representation of a missing VP sentence makes it relatively acceptable, even though the sentence is ungrammatical?

2.1 Why is a VP omitted?

2.1.1 Structural forgetting

The classic structural forgetting account is proposed by Gibson & Thomas (1999). While they do not explicitly state what representations are built, Gibson & Thomas appear to assume that a sentence can be encoded, at a minimum, as a set of NPs, VPs, and so on. They also appear to assume a process that combines constituents to produce argument-predicate relations, although they say little about how the combinatorial process works for relativization. The parser predicts upcoming constituents based on the constituents observed so far, but can forget certain predictions under high memory load.

[2] We therefore set aside accounts attributing the illusion to prosodic factors (e.g. Fodor et al. 2017). Briefly, these accounts observe that grammatical DCE sentences typically have an unnatural prosodic structure that is unbalanced or list-like. It is argued that one way to improve the sentence's prosody is by omitting a constituent, like a VP. While it helps explain why grammatical DCE sentences are unacceptable, it does not address how the ungrammatical sentence is perceived as acceptable, nor why the illusion occurs with a missing middle VP, nor why the illusion varies across languages.

For illustration, we describe how the parser might process a DCE sentence like (1). The first NP, *the novel*, triggers a prediction for a verb, required to form a complete sentence. The second NP, *the horror author*, signals to the parser the presence of an object relative clause, triggering a prediction for a VP – specifically, a verb and an empty category. The third NP, *the publishing company*, also triggers the prediction for a VP, consisting of another verb and empty category. These predictions are costly to maintain, and the parser forgets the prediction with the highest memory cost in order to reduce memory load.

In Gibson & Thomas's theory of memory load, prediction costs depend on structural position and the number of discourse referents processed since a prediction was made. Consider the prediction costs at the point at which all three NPs have just been processed. NP1's prediction is stipulated to be costless, as it is associated with the main clause. Between NP2 and NP3, NP2's prediction is more costly, because after it was made, a further discourse referent (NP3) was introduced. No equivalent discourse referent exists for NP3. As a result, NP2's prediction gets forgotten.[3]

Gibson & Thomas also provide experimental evidence for this claim. They systematically omit each of the three VPs from a DCE sentence (3a) to produce the missing VP conditions in (3b-d). The logic of the manipulation goes as follows. If the forgotten VP prediction is the one associated with NP2 *the horror author*, then in (3b) the VP *had typed quickly* should get paired with the NP3 *the publishing company*, with this NP as the subject. Since this pairing is implausible, (3b) should be rated worse. In contrast, for (3c), the same hypothesis predicts that *the publishing company* should get interpreted as the subject of the VP *had recently fired*, while NP1 *the novel* should be interpreted as the subject of the VP *was banned by the local library*. These pairings are plausible, so (3c) should be more acceptable.

As predicted, sentences like (3c) show the missing VP illusion; sentences like (3b) and (3d) received worse ratings. To be clear, the degree to which (3c) is more acceptable than their grammatical counterparts in (3a) varies somewhat: Gibson & Thomas did not find a significant difference, but Frank & Ernst (2019) did. Regardless, the fact remains that both studies show the ungrammatical (3c) to be no worse than (3a) --- a hallmark of a grammaticality illusion.

(3) a. Grammatical DCE
[$_{NP1}$ The novel] that [$_{NP2}$ the horror author] who [$_{NP3}$ the publishing company] [$_{VP1}$ had recently fired] [$_{VP2}$ had typed quickly] [$_{VP3}$ was banned by the local library].

b. Missing VP1
[$_{NP1}$ The novel] that [$_{NP2}$ the horror author] who [$_{NP3}$ the publishing company] ---- [$_{VP2}$ had typed quickly] [$_{VP3}$ was banned by the local library].

c. Missing VP2
[$_{NP1}$ The novel] that [$_{NP2}$ the horror author] who [$_{NP3}$ the publishing company] [$_{VP1}$ had recently fired] ---- [$_{VP3}$ was banned by the local library].

d. Missing VP3
[$_{NP1}$ The novel] that [$_{NP2}$ the horror author] who [$_{NP3}$ the publishing company] [$_{VP1}$ had recently fired] [$_{VP2}$ had typed quickly] ----.

[3] Gibson & Thomas also mention the possibility that it is the "word-string" whose predictions are costly that gets forgotten (pp. 241—242). In other words, the parser forgets the second NP *the horror author*, along with its predictions. As we understand it, this alternative analysis predicts that the relative clause that follows this NP, consisting of the third NP and first VP, should be headless. This does not seem to be the case: the relative clause quite clearly still gets interpreted as modifying the second NP.

The structural forgetting hypothesis attributes the missing VP effect to general limitations on cognitive (specifically, memory) resources. In this regard, the hypothesis also has conceptual parallels with Christiansen & Chater's (1999) account, in which the illusion is argued to reflect architectural constraints of the parser (also see Christiansen & MacDonald 2009). However, Christiansen & Chater's connectionist account does not go as far as to predict that it is the middle VP that goes missing. This is because it represents sentences as flat sequences of nouns and verbs, and so does not predict how words are related to each other.

2.1.2 Language experience

Because it explains the missing VP effect as the result of general memory limitations, the structural forgetting hypothesis further predicts that all languages should show a missing VP illusion. However, experimental evidence suggests that this is not the case. String-equivalent German and Dutch sentences do not elicit the illusion, as shown by reading time studies (Vasishth et al. 2010; Frank et al. 2016; but see Häussler & Bader 2015 and Bader 2016 for data on other kinds of German DCE sentences, which we will set aside here) and offline acceptability and comprehensibility rating studies (Frank & Ernst 2019).

Vasishth, Frank, and colleagues suggest that these cross-linguistic differences reflect differences in language experience. They point out that German and Dutch consistently have verb-final embedded clauses, unlike English. As a result, in German and Dutch, verbs are more likely to appear relatively far from their subjects. Sequences of verbs are also more common. Consequently, speakers of these languages are better at parsing grammatical DCE sentences and detecting the ill-formedness of missing VP sentences. Both types of sentences feature sequences of two or three verbs, such that some of the verbs appear some distance away from their subjects.[4]

This hypothesis has been fleshed out in two ways. For instance, Vasishth et al. (2010) integrate it with the structural forgetting hypothesis. They suggest that German and Dutch speakers' experience with processing verb-final structures can condition their working memory, so predicted verbs have "more robust memory representations" (p. 558) in these languages than in English. Consequently, German and Dutch speakers are less likely to forget predictions, even under high memory load. Similar intuitions are implemented in the computational models of Engelmann & Vasishth (2009), Futrell & Levy (2017), and Futrell et al. (2020).

The second option is a shallower theory that avoids making similar commitments about memory representations, mentioned by Frank et al. (2016) and Frank & Ernst (2019). This theory assumes implicitly that sentences can be represented linearly as sequences of nouns and verbs (see also Christiansen & Chater 1999; Christiansen & MacDonald 2009). Because sequences of three consecutive verbs are more common in German and Dutch than in English, they are read more easily and rated as more acceptable, even in DCE contexts. However, as Frank & Ernst and Futrell et al. note, this approach results in a weaker theory of the illusion: it predicts that an English DCE sentence should become more acceptable as long as any of the VPs is omitted, a prediction inconsistent with Gibson & Thomas's results.

[4] One question that arises is just what kind of experience German and Dutch speakers have with these kinds of sentences. As far as we know, there is no published quantitative study of this question, although Frank et al. (2016:568) assert that three-verb sequences are more common in German and Dutch.

2.1.3 Interference

Häussler & Bader (2015) and Bader (2016) present alternative proposals that appeal to interference and primacy and recency effects, all well-established concepts in the memory retrieval and sentence processing literature (see also Gibson & Thomas 1999:242-244). In both accounts, the parser attaches words into a syntactic representation as they appear. Both accounts also assume that recency effects cause the parser to correctly attach the first verb to the VP in the lower relative clause: the parser has just processed the third NP as the subject of the lower relative clause, so the relative clause (or its VP) is still highly activated.

At that point, the representation contains two empty positions for verbs, one in the main clause and one in the higher relative clause. Both positions compete for the attachment for the next verb. Häussler & Bader (2015) observe that the main clause was the first clause created by the parser. It therefore benefits from a primacy effect, making it more likely that the second verb is incorrectly attached there, leaving the relative clause without a verb.

Bader (2016) presents a structural hypothesis that can explain why the strength of the illusion differs between English and German. German has verb-second word order in main clauses, so the two attachment sites differ syntactically: the relative clause site is inside a verb-final VP, while the main clause site is a verb-second position in CP. The German parser can therefore discriminate between the two sites. In contrast, English does not have verb-second word order. Both attachment sites are inside VPs, so they are more easily confused, making wrong attachment to the main clause site more likely. Although Bader does not go into details, this proposal may be compatible with a cue-based retrieval system, in which structural features like "VP" and "CP" are part of the set of retrieval cues associated with a verb.

To explain why German speakers can successfully go on to attach the third verb to the main clause (instead of overlooking or forgetting the need for a third verb by the time it appears), Bader appears to suggest that the distinctiveness of the main clause attachment site in German makes it easier to detect when it is missing a verb. He also suggests that processing load might play a role. In these sentences, the subject is structurally high; he argues that this position is associated with a lower processing load, freeing up resources that improve parsing accuracy. For thoroughness, we also note, following Häussler & Bader (2015), that primacy effects might provide yet another explanation: the main clause attachment site was created early, and so is less easily overlooked.

2.2 What makes English missing VP sentences acceptable?

In contrast to the debate on why a VP gets omitted, there is less discussion about why ungrammatical missing VP sentences can be perceived as more acceptable. Implicit in existing accounts is the idea that the parser comes to develop certain expectations about the VPs in these sentences. These expectations are then borne out, producing a sense of completeness.

We can distinguish between two variants of this idea. The first variant can be found in an account like Gibson & Thomas's and computational models of the illusion, in which sentences are represented as a sequence of constituents with minimal syntactic structure. Memory limitations cause the parser to expect only two VPs, not three. Since the parser observes only two VPs, these predictions are borne out, producing a sense of completeness and acceptability.

In contrast, interference accounts hold that the parser builds a fuller hierarchical structure with nodes for upcoming VPs. It is implicitly assumed that the parser expects to fill all nodes, even though the parser ends up leaving the intermediate VP node empty and fails to notice it. Regardless

of why exactly the empty node goes unnoticed, the result is that the parser mistakenly concludes that all three VP nodes have been filled and the sentence is complete.

Here, we offer an alternative, non-structural perspective that does not rely on the parser's expectations about certain nodes. Our hypothesis is based on the premise that a core goal in parsing is to establish thematic relations between the referents and predicates denoted by the NPs and verbs in a sentence: who did what to whom, so to speak (see Frazier 1985; Pritchett 1988; Boland et al. 1995; Gibson 1998; Aoshima et al. 2004; among many others, for similar proposals). Setting aside what processes cause VPs to be omitted, we suggest that missing VP sentences are acceptable because thematic relations can be successfully built between all remaining arguments and predicates. Because no argument or predicate is thematically "orphaned," speakers perceive a sense of completeness, which translates into an illusion of grammaticality.

For illustration, consider the missing VP sentence in (3c): *The novel that the horror author who the publishing company had recently fired was banned by the local library*. In this sentence, all three NPs (more precisely, the arguments they represent) are related to some VP predicate: NP1 *the novel* is related to VP2 *was banned by the local library*, while NP2 *the horror author* and NP3 *the publishing company* are related to VP1 *had recently fired*. Conversely, the two predicates in VP1 and VP2 collectively require three arguments: two for *had recently fired*, and one for *was banned by the local library*. These requirements are satisfied by the three NPs. It is easy to see from Figure 1 that every argument and predicate is connected to something else. (For comparison purposes, we depict thematic relations for a grammatical DCE sentence in Figure 2.)

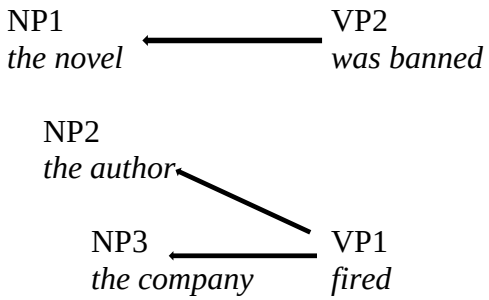


Figure 1: Thematic relations (depicted with arrows) in missing VP sentence (3c) [$_{NP1}$ The novel] that [$_{NP2}$ the horror author] who [$_{NP3}$ the publishing company] [$_{VP1}$ had recently fired] [$_{VP2}$ was banned by the local library].

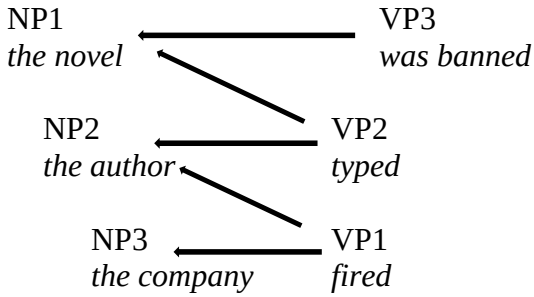


Figure 2: Thematic relations in grammatical DCE sentence (3a) [$_{NP1}$ The novel] that [$_{NP2}$ the horror author] who [$_{NP3}$ the publishing company] [$_{VP1}$ had recently fired] [$_{VP2}$ had typed quickly] [$_{VP3}$ was banned by the local library].

An important question here is what kind of representation is implicated in this thematic relations account. It is difficult to say definitively that the relevant representation must be syntactic and that thematic relations are read off from that structure. In a grammatical DCE sentence, NP1 and NP2 should each be syntactically related to two verbs: the verb of the relative clause modifying them and the verb that they are the subjects of, as reflected by the thematic relations in Figure 2. It is unclear what happens to these syntactic relations in a missing VP sentence, where there is only one verb each for NP1 and NP2. In fact, as Figure 1 shows, a missing VP sentence appears to consist of two disjoint sets of thematic relations – one connecting NP1 and VP2, and another connecting NP2, NP3, and VP1 – thus implying two disjoint sets of syntactic relations.

For this reason, we do not rule out the possibility that in addition to a syntactic representation, there is a shallower representation available that allows the parser to track what thematic relations exist between the arguments and predicates denoted by NPs and verbs. We remain agnostic about the specifics of this shallower representation, because of the wide range of options available. Thematic relations could be established directly between NPs and verbs, if they can be represented as discrete subtrees, or between arguments and predicates within a semantic representation. Both approaches may be amenable to a cue-based retrieval analysis (Lewis & Vasishth 2005; Brasoveanu & Dotlačil 2020).

Of course, to the extent that shallow representations are involved, this suggestion seems inconsistent with a body of psycholinguistic research. There is substantial evidence that the parser incrementally builds a rich, connected syntactic representation, as revealed by the finding that the real-time building of various dependencies is sensitive to syntactic constraints (for filler-gap dependencies, see Stowe 1986; Traxler & Pickering 1996; Phillips 2006; Wagers & Phillips 2009; etc.; for anaphoric dependencies, see, e.g., Sturt 2003; Kazanina et al. 2007; Aoshima et al. 2009).

We suggest that this inconsistency can be resolved if we consider the exceptional syntactic complexity and memory load associated with DCE sentences. In English, for example, at the point when the third NP subject is processed, the parser must maintain in memory all three subjects and the predictions associated with each of them (Miller & Isard 1964; Resnik 1992; Gibson & Thomas 1999, etc.). Under such circumstances, the parser might find it difficult to build and maintain a connected structure for interpretation. Consequently, it might rely on a shallower interpretive strategy, directly tracking how arguments and predicates are linked to each other, so that it can produce a passable interpretation of the sentence.

As presented, the thematic relations account, as well as the structural forgetting and interference accounts, predict missing VP sentences to be highly acceptable. A challenge for this prediction is that these sentences actually receive relatively low ratings (Gibson & Thomas 1999 and Experiment 3 below; Frank & Ernst 2019). The low ratings might simply reflect the length and structural complexity of these sentences, although they might also be seen as evidence that speakers maintain a syntactic representation of these sentences. The low ratings can then be attributed to a percept of syntactic ill-formedness that overshadows the sense of completeness.

Another possibility is that the processes above are less deterministic than depicted, so missing VP sentences do not consistently elicit a sense of completeness. In an interference approach, the parser might sometimes attach the second verb correctly to the intermediate VP node, leaving the main clause VP node empty. If empty main clause VP nodes are easier to detect, as hypothesized in Section 2.1.3, the structure would be perceived as incomplete and unacceptable. From a thematic relations perspective, leaving this node empty means that the first NP is "orphaned," reducing acceptability. Alternatively, the memory load on the parser might cause it to sometimes fail to track and establish thematic relations between all arguments and predicates.

Overall, the thematic relations hypothesis makes very similar predictions as existing accounts about missing VP sentences. However, a key difference is that it does not require the parser to overlook the fact that the second NP in these sentences is missing a VP. Rather, the absence of a VP is essentially offset by the fact that this NP can still be linked to another predicate. One way to test this hypothesis, then, would be to consider a double center-embedding scenario where this link is unavailable, so the second NP (or its equivalent) runs the risk of being "orphaned." Mandarin Chinese provides just such a test case.

3 The present study

As we show below, the word order properties of Mandarin conspire to produce DCE sentences in which a sequence of three transitive verbs is followed by a sequence of NPs. Mandarin also appears to exhibit a missing NP illusion. At first glance, in these missing NP sentences, the second transitive verb – the equivalent of the second NP in English missing VP sentences – seems to be thematically orphaned. As we elaborate in this section, it is unclear which NPs can serve as its arguments. We further highlight challenges in adapting existing hypotheses about missing VP illusions to account for the missing NP illusion.

We present a series of experiments to substantiate these points. Experiment 1 demonstrates the missing NP illusion and determines how the transitive verbs and the NPs that follow them are thematically related. Experiment 2 strengthens the claim for the illusion, by showing that these sentences are unlikely to have an alternative grammatical parse with only one relative clause.

An unexpected finding of Experiment 1 is that the final NP in missing NP sentences is interpreted as an argument of all three verbs, so no verb is truly orphaned. This interpretation diverges sharply from English, where the final VP (the analog of the Mandarin final NP) is interpreted only with the first NP. To ensure the validity of this comparison, we replicate Gibson & Thomas's experiment in Experiment 3, which demonstrates the missing VP effect and provides evidence for which NPs and VPs are thematically related.

Finally, we argue that the language experience account, intended as an account of cross-linguistic variation of the missing VP illusion, does not predict the missing NP illusion. We do so with a computational simulation (Experiment 4), adapting Futrell and colleagues' model of the language experience account (Futrell et al. 2020; also Futrell & Levy 2017).

3.1 Center-embedding in Mandarin

Mandarin has head-final relative clauses in which a clause-final particle *de* is followed by the head noun. Suppose that a noun NP3 is modified by a subject relative clause with an object NP2. This complex noun phrase has [V-NP2-*de*]-NP3 word order. Since Mandarin is a Subject-Verb-Object language, a sentence where this complex NP is the object has Subject-V1-[V2-NP2-*de*]-NP3 word order. Inserting another subject relative clause to modify NP2 yields a DCE sentence with Subject-V1-[V2-[V3-NP1-*de*]-NP2-*de*]-NP3 word order (4).

(4) DCE sentence

Zongli	zhengzai	[$_{V1}$	jiejian]	ceng	[$_{V2}$	zebeiguo]	gang	[$_{V3}$	shenli]
prime.minister	PROG		meet	previously		rebuked	recently		hear
[$_{NP1}$	tanwu-an]	bujiu	de	[$_{NP2}$	faguan]	haojici	de	[$_{NP3}$	buzhang].
	corruption-case	not.long.ago	DE		judge	several.times	DE		minister

'The prime minister is meeting the minister who previously rebuked several times the judge who just heard the corruption case not long ago.'

A missing NP sentence can be constructed by omitting one of the *de*s and NPs (5), by analogy with missing VP sentences.

(5) Missing NP sentence
Zongli zhengzai [$_{V1}$ jiejian] ceng [$_{V2}$ zebeiguo] gang [$_{V3}$ shenli]
prime.minister PROG meet previously rebuked recently hear
[$_{NP1}$ tanwu-an] bujiu de [$_{NP2}$ faguan].
corruption-case not.long.ago DE judge
'The prime minister is meeting ---- who previously rebuked several times the judge who just heard the corruption case not long ago.'

As the English "translation" of (5) shows, a missing NP sentence is clearly ungrammatical: the three transitive verbs each require an object, but there are only two NPs in object positions. However, impressionistically, such a sentence is as acceptable as its grammatical counterpart (4), if not more so – it is this relative acceptability that we refer to as the missing NP illusion.

3.2 Ruling out null object or headless relative clause analyses

One might wonder whether we might have been too hasty in describing missing NP sentences as ungrammatical. In Mandarin, the presence of appropriate discourse antecedents licenses null objects (6a) and "headless" relative clauses where the head NP is elided (6b). However, there are two reasons why it is inappropriate to assimilate missing NP sentences with either construction. First, both constructions elicit a specific percept of ill-formedness when presented out of context, when no antecedent is available (e.g. Huang 1984; Li 2014). In contrast, missing NP sentences are relatively acceptable even out of context.

(6) Null object and headless relative clauses infelicitous without appropriate context
a. #Tamen zebeiguo __ haojici.
they rebuked several.times
'They rebuked [them/her/…] (several times).'
b. #[$_{RC}$ Wo zuotian mai de] __ mei name gui.
I yesterday buy DE not that expensive
'The [food/clothes/…] I bought yesterday was not that expensive.'

Second, missing NP sentences can be distinguished syntactically from both null object and headless relative clause constructions. In null object sentences, verbs usually appear at the end of a clause (embedded or otherwise), unless there happens to be adverbial material present (6a). In contrast, in missing NP sentences, none of the verbs appear in such a position. Both V1 and V2 precede another verb that appears to start a subject relative clause, while V3 immediately precedes a direct object NP.

Missing NP sentences also do not have the syntactic profile of headless relative clause sentences. In missing NP sentences, what goes missing is an NP and a *de*. In contrast, only an NP is omitted in headless relative clauses like (6b).

3.3 Challenges for existing accounts

Abstractly, missing NP sentences resemble missing VP sentences: missing NP sentences feature a sequence of three verbs and two NPs, while missing VP sentences feature a sequence of three NPs and two VPs. However, it turns out to be challenging to adapt the hypotheses proposed for the missing VP illusion to account for the missing NP illusion. Both structural forgetting and interference accounts predict very different thematic relations for English missing VP sentences and Mandarin missing NP sentences, raising questions about why the Mandarin sentences should be relatively acceptable.

We first summarize the intuition behind this argument. In English missing VP sentences, the parser effectively overlooks the absence of a VP for the second NP. Despite the absence of a VP, the parser is still able to establish thematic relations between the remaining NPs and VPs.

By analogy, in Mandarin missing NP sentences, the parser should have also forgotten or overlooked the predictions for an NP associated with the second verb. Here, however, the absence of an NP should cause a more severe disruption in terms of thematic relations, which should reduce acceptability. The three verbs in missing NP sentences are transitive, each requiring two arguments. However, including the subject of the main clause, there are only three NPs. Intuitively, there are not enough NPs. We argue below that there is a distinct possibility that V2 would be "orphaned" thematically, without agent and theme arguments.

More specifically, let us consider these sentences from an incremental parsing perspective, following the structural forgetting and interference accounts. (We have simplified the following description, for exposition purposes. We refer interested readers to Jäger et al. 2015 for an overview of the local ambiguities that a Mandarin parser must navigate in parsing a relative clause.)

Suppose that a missing NP sentence like (5) is parsed in the following manner. First, the parser observes an NP, *prime minister*, and a transitive V1 *meet*. It interprets the NP as V1's subject and agent argument. Observing V1 triggers a prediction for an NP object, its theme argument.

Next, when a transitive V2 *rebuked* is seen rather than V1's object, V2 is recognized as the main verb of a subordinate clause modifying V1's object. Being transitive, V2 triggers predictions for an NP object and a *de*. To the extent that the parser analyzes this subordinate clause as a subject relative clause, it should also expect the NP appearing after *de* to be interpreted as V2's agent.

Lastly, when a transitive V3 *hear* is seen rather than V2's object, V3 is recognized as the main verb of a subordinate clause modifying V2's object. The parser again should predict an NP object (V3's theme) and *de*. Again, to the extent that the parser analyzes this subordinate clause as a subject relative clause, it should expect the NP appearing after *de* to be interpreted as V3's agent.[6]

Both structural forgetting and interference hypotheses claim that the parser should forget or overlook the predictions associated with V2 (but not V2 itself). Consequently, the parser should pair the first post-verb NP and *de* with V3, and the second NP with V1. To the extent that the

[6] As mentioned, the actual parsing process is probably more complicated (see Lin & Bever 2006; 2011; Vasishth et al. 2013; Jäger et al. 2015). As a preview, consider what happens when the parser first encounters V2 in our example. It is possible that the parser analyzes the verb as the main predicate of something other than a subject relative clause. For instance, because Chinese allows null arguments, the parser could assume that the verb's subject is null, especially if there is an appropriate antecedent licensing a null subject in the discourse context. In such an analysis, V2 might be the predicate of the complement clause of V1's object, or the predicate of an object relative clause. Alternatively, with the right V1, the parser might assume that V1 can take a clause or VP-like complement in which V2 is the main verb.

parser analyzes V3 as a subject relative clause, the second NP should also end up getting interpreted as V3's agent.

Figure 3 sums up the thematic relations built in this scenario. The three NPs – the main clause subject, NP1, and NP2 – are all linked to verbs. As for the three verbs, only two have arguments: V1 and V3; V2 is "orphaned." If missing VP sentences are acceptable because the parser can build thematic relations between predicates and arguments, one might predict these sentences to be less acceptable than they are reported to be. (For comparison, Figure 4 depicts thematic relations for a grammatical DCE sentence.)

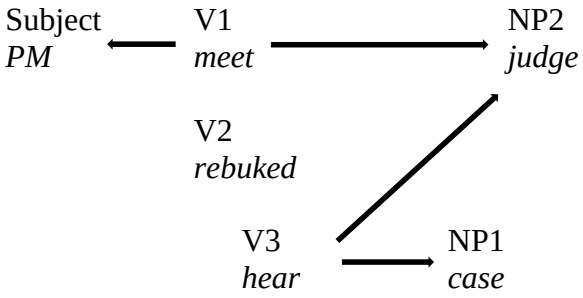


Figure 3: Hypothetical thematic relations in Mandarin missing NP sentence (5).

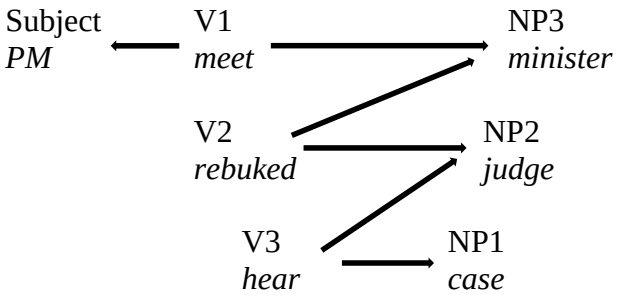


Figure 4: Thematic relations in grammatical Mandarin DCE sentence (4).

One could certainly accommodate the missing NP illusion with a "shallower" approach in which sentences are represented as sequences of syntactic categories, without deeper commitments about thematic relations. As mentioned in Section 2.1.2, in such an approach, the parser predicts lexical items based on what it has seen so far. Upon processing V3, the parser might expect to see only a noun, then a *de*, followed by another noun. These predictions are borne out in missing NP sentences, which explains why the sentences are acceptable. However, this "shallower" approach has little to say about why the two VPs in a missing VP sentence are systematically interpreted with the first and third NPs as their subjects. Consequently, this approach provides no insight into how these interpretations contribute to percepts of acceptability.

Alternatively, one might wonder if the "orphaning" of V2 in Figure 3/(5) is only superficial. Since Mandarin allows null arguments, perhaps the parser can supply null arguments for V2 from the discourse context or world knowledge, so V2 can satisfy its argument structure needs without being linked to other NPs. If so, it should be easy to come up with a grammatical version of (5) that has overt arguments for V2 and is consistent with the thematic relations in Figure 3. But there is no acceptable way to add two more overt NPs for V2 to a missing NP sentence like (5).

Taking a step back, one might find it unsurprising that Mandarin poses challenges for the structural forgetting hypothesis. After all, this hypothesis faces empirical problems elsewhere: it

incorrectly predicts that German and Dutch should exhibit the missing VP illusion in the same way as English. However, supplementing the structural forgetting hypothesis with the language experience hypothesis, as suggested by Vasishth et al. (2010) for German, does not improve matters for Mandarin. In fact, doing so might yield the prediction that Mandarin should lack such an illusion. (Similar remarks hold for alternative implementations of this hypothesis, as presented in Engelmann & Vasishth 2009, Frank et al. 2016; Futrell et al. 2020.)

The root of this problem lies in the syntax of Mandarin noun phrases: they are uniformly head-final; all complements and modifiers are marked overtly with *de*. Consequently, sequences of nouns, whether interleaved with *de* or not, are common in Mandarin: they appear in compound (7a) and possessive structures (7b). They also appear when a noun appears with a subordinate clause that contains an object, regardless of whether the clause is a complement (7c) or relative clause (7d). In fact, subject relative clauses, which correlate with noun sequences (7d), are reported to occur more frequently than object relative clauses (Hsiao & Gibson 2003; Vasishth et al. 2013).

(7) Mandarin noun sequences (NPs in sequences marked with square brackets)

a. [$_{NP}$ daxue] [$_{NP}$ xiaozhang] [$_{NP}$ bangongshi]
university president office
'office of the university president'

b. [$_{NP}$ kexuejia] (de) [$_{NP}$ shiyan]
scientist DE experiment
'(the) scientist's experiment'

c. [buzhang zebei [$_{NP}$ faguan]] de [$_{NP}$ xiaoxi]
minister rebuke judge DE news
'the news that the minister rebuked the judge'

d. [zebei [$_{NP}$ faguan]] de [$_{NP}$ buzhang]
rebuke judge DE minister
'the minister who rebuked the judge'

Mandarin speakers should therefore be familiar with processing sequences of nouns, just as German and Dutch speakers are familiar with processing sequences of verbs. In addition, given Mandarin's VO word order and head-final NPs, it is not unusual for a verb to be separated from the head of its object NP by a modifier, such as a possessor or a relative clause. The net effect is that Mandarin speakers should find grammatical DCE sentences, which feature long verb-object dependencies and sequences of nouns, easier to process than missing NP sentences.

Of course, the points here assume that the missing NP illusion exists in Mandarin and Mandarin speakers do parse and interpret missing NP sentences as depicted in Figure 3, where V2 is thematically orphaned. In the next section, we report an experiment designed to validate these assumptions. As a preview, we find that missing NP sentences are relatively acceptable, but no verb is orphaned. Instead, speakers can and do interpret V2 as taking NP2 as an argument.

4 Experiment 1

4.1 Materials

24 DCE sentences were constructed in simplified Chinese, each with six versions corresponding to different experimental conditions. Two of these conditions were grammatical

DCE sentences and "plausible" missing NP sentences, designed to test for the missing NP illusion. The remaining four conditions were "implausible" missing NP sentences, designed to determine which verbs and NPs appearing after them are thematically connected in missing NP illusions. We distributed these 24 sentences into six lists using a Latin Square design, so that each list contained four sentences for each condition, and no two sentences in each list were variants of each other.

In addition, 48 fillers were constructed. These fillers were designed to be of a similar length and syntactically complex, featuring subordinate clauses, conjoined clauses, and non-canonical word order. 36 of the fillers were acceptable sentences and 12 were unacceptable ones. This distribution was chosen so that participants would read an equal number of acceptable and unacceptable sentences in the combined experiment, assuming that participants find all target sentences less acceptable than the acceptable fillers.

In all target sentences, a duration or frequency adverb appeared between the NP objects and the clause-final particle *de*. In (4), repeated below as (8), the duration adverb *bujiu* "not long ago" appears between NP1 *tanwu-an* "corruption case" and *de*, and the frequency adverb *haojici* "several times" between NP2 *faguan* "judge" and *de*.

(8) Mandarin DCE sentence

Zongli	zhengzai	[$_{V1}$ jiejian]	ceng	[$_{V2}$ zebeiguo]	gang	[$_{V3}$ shenli]		
prime.minister	PROG	meet	previously	rebuked	recently	hear		
[$_{NP1}$ tanwu-an]	bujiu	de	[$_{NP2}$ faguan]	haojici	de	[$_{NP3}$ buzhang].		
corruption-case	not.long.ago	DE	judge	several.times	DE	minister		

'The prime minister is meeting the minister who previously rebuked several times the judge who just heard the corruption case not long ago.'

These adverbs served to disambiguate the reading of *de*. Without an adverb, *de* in this position – between two nouns like *tanwu-an* "corruption case" and *faguan* "judge" – is in principle ambiguous. *De* can mark the end of a relative clause, in which case *tanwu-an* is inside the relative clause and *faguan* is the head of the relative clause. Alternatively, *de* can be understood as a possessive marker connecting two NPs, so "corruption case" and "judge" form a single complex NP (9). The adverb's presence blocks the possession reading.

(9)

tanwu-an	de	faguan
corruption-case	DE	judge

'(the) corruption case's judge'

DCE sentences like (8) were contrasted with "plausible" missing NP variants like (10), which end prematurely at NP2. These variants were plausible in the respect that all likely verb-argument relations were semantically plausible. For instance, (10) is plausible in that NP2, *faguan* "judge" is a plausible theme argument for either V1, *jiejian* "meet," or V2, *zebeiguo* "rebuked," and a plausible agent of V3 *shenli* "hear." Of course, they were not plausible in the sense that they have a fully coherent meaning, since the sentences were incomplete.

(10) Mandarin plausible missing NP sentence

Zongli	zhengzai	[$_{V1}$ jiejian]	ceng	[$_{V2}$ zebeiguo]	gang	[$_{V3}$ shenli]		
prime.minister	PROG	meet	previously	rebuked	recently	hear		
[$_{NP1}$ tanwu-an]	bujiu	de	[$_{NP2}$ faguan].					

corruption-case not.long.ago DE judge
A possible translation (if there is one): 'The prime minister is meeting ---- previously rebuked the judge who just heard the corruption case not long ago.'

We chose verbs, aspect markers, and adverbs so as to maximize the possibility that participants parse a missing NP sentence as doubly center-embedded. First, we tried to pick verbs with a clear NP complement bias, avoiding verbs that can also readily appear with only a clause-like complement. This meant excluding verbs that describe speech, cognition, or desire, such as *xihuan* "like." The goal was to prevent participants from assigning an unintended, alternative grammatical parse with only a single level of embedding (11b); we expand on this point in Experiment 2. To identify these verbs, we relied on the intuitions of the first author, a native speaker of Mandarin. We also verified these intuitions with the argument structure listings of these verbs in the Chinese Proposition Bank (Xue & Palmer 2008).

(11) Zongli [$_{V1}$ **xihuan**] [$_{V2}$ zebei] gang [$_{V3}$ shenli] [$_{NP1}$ tanwu-an]
prime.minister like rebuke recently hear corruption-case
bujiu de [$_{NP2}$ faguan].
not.long.ago DE judge
a. Intended missing NP parse (if there is one): 'The prime minister likes ---- rebuked the judge who just heard the corruption case not long ago.'
b. Unintended grammatical parse: 'The prime minister likes to rebuke the judge who just heard the corruption case not long ago.'

Second, verbs within a sentence also had distinct lexical semantics, and where possible, distinct grammatical aspect morphology (such as experiential *-guo* or perfective *-le*, glossed using the English past tense). As a generalization, two verbs can be conjoined in Mandarin without an overt marker when the verbs have similar semantics and are in the same grammatical aspect. By controlling for these factors, we discouraged participants from coercing a reading of a missing NP sentence where two of the three verbs, such as V1 "criticize" and V2 "rebuke" in (12), are conjoined with each other. A conjunction analysis would produce a grammatical sentence with only one level of embedding, as illustrated by the English translation in (12b).

(12) Zongli [$_{V1}$ piping] [$_{V2}$ zebei] gang [$_{V3}$ shenli] [$_{NP1}$ tanwu-an]
prime.minister criticize rebuked recently hear corruption-case
bujiu de [$_{NP2}$ faguan].
not.long.ago DE judge
a. Intended missing NP parse (if there is one): 'The prime minister criticized ---- rebuked the judge who just heard the corruption case not long ago.'
b. Unintended grammatical parse: ''The prime minister criticizes and rebukes the judge who just heard the corruption case not long ago.'

Third, we added past-oriented temporal adverbs like *ceng(jing)* "previously," *yi(jing)* "already," *gang* "just recently," *zhiqian* "in the past" after the first verb and after the second verb, as suggested by an anonymous reviewer. These temporal adverbs seem to encourage these verbs to be parsed as taking a complex NP object, rather than a complement clause.

Four other conditions were created by modifying the plausible missing NP condition, so that the NPs following the verbs would be perceived as inappropriate arguments for plausibility and/or animacy reasons, depending on which of the three verbs the NP was interpreted with. These manipulations parallel Gibson & Thomas's manipulations for their English missing VP experiment (3), which were also intended to detect which nouns and verbs are interpreted together.

Three of these conditions manipulated verbs to alter the plausibility of NP2 as an argument. In the first condition, V1 (and associated aspect/modal markers) was chosen so that NP2 would be an unlikely or inappropriate theme argument (13a). While we do not translate sentences like (13a) because doing so would presuppose a parse, the intuition for this condition is as follows: if speakers interpret NP2 *faguan* "judge" as the theme argument of V1 *chongzu* "reorganize," they should assign a lower acceptability rating to the sentence, since one reorganizes institutions, not judges.

Likewise, in the second condition, V2 was chosen so that NP2 would be an inappropriate theme argument (13b), if NP2 were so interpreted. In the third condition, V3, its object (NP1), and adverb were chosen so that NP2 would be an unlikely or inappropriate agent (13c): judges do not lead national basketball teams. For ease of reference, we call these the "implausible V1 object," "implausible V2 object," and "implausible V3 subject" conditions.

(13) Manipulations involving verbs and NP2

a. Implausible V1 object

Zongli	jiang	[$_{V1}$ **chongzu**]	ceng	[$_{V2}$ zebeiguo]	gang	[$_{V3}$ shenli]	
prime.minister	FUT	organize	previously	rebuked	recently	hear	
[$_{NP1}$ tanwu-an]	bujiu	de	[$_{NP2}$ **faguan**].				
corruption-case	not.long.ago	DE	judge				

b. Implausible V2 object

Zongli	zhengzai	[$_{V1}$ jiejian]	zuijin	[$_{V2}$ **chongzule**]	gang	[$_{V3}$ shenli]	
prime.minister	PROG	meet	recently	reorganized	recently	hear	
[$_{NP1}$ tanwu-an]	bujiu	de	[$_{NP2}$ **faguan**].				
corruption-case	not.long.ago	DE	judge				

c. Implausible V3 subject

Zongli	zhengzai	[$_{V1}$ jiejian]	ceng	[$_{V2}$ zebeiguo]	yijing	[$_{V3}$ **shuailing**]	
prime.minister	PROG	meet	previously	rebuked	already	lead	
[$_{NP1}$ **guojia**	**lanqiu-dui**]	bujiu	de	[$_{NP2}$ **faguan**].			
national	basketball-team	not.long.ago	DE	judge			

The fourth and last condition involved NP1. NP1 was chosen so that it would be an implausible theme object of V3, the verb that immediately precedes it (14). We refer to this as the "implausible V3 object" condition.

(14) Implausible V3 object (*shucai* "vegetables" is an implausible object for V3 *shenli* "hear")

Zongli	zhengzai	[$_{V1}$ jiejian]	ceng	[$_{V2}$ zebeiguo]	gang	[$_{V3}$ **shenli**]	
prime.minister	PROG	meet	previously	rebuked	recently	hear	
[$_{NP1}$ **shucai**]	bujiu	de	[$_{NP2}$ faguan].				
vegetables	not.long.ago	DE	judge				

Obviously, there are other implausible pairings that we could have created. We chose the above four conditions, however, because there is prior reason to think that speakers might interpret

the NPs and verbs as standing in these thematic relations. It is plausible that NP2 gets interpreted as V1's object, by analogy to English missing VP sentences, where the first NP is interpreted as the second VP's subject. One might also expect that NP2 might get interpreted as V2's object and as V3's subject, as these are the interpretations that NP2 would receive in grammatical DCE sentences. Similarly, one would expect NP1 to be interpreted as V3's object, by analogy to grammatical DCE sentences and by the fact that NP1 appears in a canonical object position.

4.2 Participants

Participants were 60 self-identified native speakers of Mandarin Chinese, who were born in mainland China. All participants were above the age of 18. They were recruited over Prolific and compensated US$3.17 for their time, based on the time estimated necessary for completing the experiment and Prolific's recommended hourly rate of US$9.50.

We recruited participants by first asking them to complete a ten-question multiple choice screening test on Prolific, for which they were compensated US$0.81 for their time (again based on Prolific's recommended rate). In addition to ensuring that participants could read simplified Chinese proficiently, the test also checked for familiarity with Mandarin syntax and vocabulary, including idiomatic expressions that non-native or heritage speakers might find difficult. We selected only participants who answered at least eight questions correctly.

4.3 Procedure

Sentences were presented using Ibex (created by Alex Drummond). Participants were instructed to rate the acceptability of each sentence with a 7-point Likert scale, where "1" was "very incoherent (*bu tongshun*), totally unacceptable" and "7" was "very coherent, totally acceptable." For the first four practice sentences, participants saw brief comments about each sentence. For example, they were instructed that a sentence that violated Mandarin word order rules or described an unlikely scenario should receive a low rating. One of these ungrammatical practice sentences was the following non-DCE sentence, which is missing an NP (15). This sentence was modeled after one of Gibson & Thomas's English practice sentences, which also featured a non-DCE sentence missing a VP. Following Gibson & Thomas, participants were told that this sentence should receive a low rating because it was not coherent, although it was not explained exactly how it was not coherent.

(15) Missing NP practice sentence
Zhe ming hushi feichang guanxin zhe ge weikou bu hao.
this CL nurse very.much concern this CL appetite not good
Intended parse (if there is one): 'This nurse is very much concerned with this ---- whose appetite is poor.' (To make the sentence complete, one might add a *de* and *bingren* "patient" after *weikou bu hao*.)

Each sentence appeared with a rating scale. Participants rated the sentence using the keyboard or by clicking on the rating scale. Even though there was no time limit, participants were also instructed to give a response as quickly as they could, to discourage them from re-reading sentences and noticing any structural anomalies.

4.4 Data analysis

The target sentences in this experiment were fairly long. To ensure that participants had taken the time to read a target sentence before rating it, we excluded responses that were provided in less than 3 seconds after the presentation of the sentence. This eliminated 2.8% of responses.

Acceptability ratings for all conditions were analyzed with a single cumulative link mixed effects model in R version 3.3.2 (R Development Core Team 2019) with the ordinal package (Christensen 2019). The model had conditions as a fixed effect and random intercepts and slopes for both participant and items.

4.5 Results

Averaged ratings for the conditions are presented in Table 1; no condition was particularly acceptable in absolute terms. Of the six types of sentences tested, the ungrammatical plausible missing NP condition received the highest ratings. Put differently, native speakers did not perceive grammatical DCE sentences to be more acceptable.

Table 1: Mean acceptability ratings for Experiment 1.
(1 = completely unacceptable, 7 = completely acceptable)

Condition	Mean acceptability rating (standard error)
Grammatical DCE (8)	3.38 (0.12)
Plausible missing NP (10)	3.66 (0.13)
Implausible V1 object (13a)	2.82 (0.12)
Implausible V2 object (13b)	2.81 (0.12)
Implausible V3 subject (13c)	3.18 (0.12)
Implausible V3 object (14)	2.54 (0.10)

Table 2 presents results of the statistical analysis, with the plausible missing NP condition as the baseline. While this condition received higher ratings than the grammatical DCE condition, the difference was not significant ($p = 0.15$). On the other hand, all four implausible missing NP conditions were less acceptable than the plausible missing NP condition (all $p < 0.01$).

Table 2: Mixed effects regression results.
(baseline: plausible missing NP condition)

Condition	Estimate	SE	z value	p value
Grammatical DCE	-0.37	0.26	-1.44	0.150
Implausible V1 object	-1.30	0.29	-4.54	<0.001
Implausible V2 object	-1.26	0.26	-4.74	<0.001
Implausible V3 subject	-0.67	0.22	-3.03	0.002
Implausible V3 object	-1.68	0.25	-6.83	<0.001

4.6 Discussion

Experiment 1 confirmed informal impressions that missing NP sentences elicit an illusion of grammaticality. The plausible missing NP condition received numerically higher ratings than its grammatical counterparts. Even though the difference was not significant, calling it a

grammatically illusion is still justified. Unlike other kinds of ungrammatical sentences, these missing NP sentences are clearly rated as no worse than their grammatical counterparts. The illusion is analogous to the missing VP illusion, despite differences in word order and thematic relations.

This result has a parallel with Gibson & Thomas 1999, which also did not find a significant contrast between missing VP sentences and their grammatical counterparts in English. It is possible, as Gibson & Thomas suggest, that the offline nature of the task let participants reread sentences and detect more easily the ungrammatical status of a missing VP sentence. For transparency, we also note that whether the numerical difference is significant depends on the random effects structure of the statistical model. When we fitted the data with a model with only random intercepts, the difference turned out to be significant, like the other four implausible conditions (all $p < 0.01$).

The lower ratings for all four implausible conditions provide a clear picture of the thematic relations in missing NP sentences. The implausible V3 object condition, for instance, confirms that the third verb is interpreted as related to the first NP.

More interesting are the contrasts for the other three conditions. Together, they suggest that speakers can relate the second post-verb NP (NP2) to all three verbs: as the theme argument (object) of V1 and V2 and as the subject (agent or experiencer) of V3. This finding is unexpected, having no analog in the missing VP illusion in English. In a missing VP sentence, the second VP is the equivalent of the Mandarin second NP, but there is no evidence that English speakers interpret this VP as taking all three NPs as its arguments, in particular, with both NP1 and NP2 as agent or experiencer arguments (subjects).

These findings do not follow easily from the structural forgetting and interference accounts. Under both accounts, the parser should have forgotten or overlooked V2's predictions for an NP object and a *de*. Since the object is interpreted as V2's theme argument, and the NP that appears immediately after *de* is interpreted as V2's agent, the forgetting or overlooking of the object and *de* should cause V2 to become thematically orphaned, without a theme or agent (see Figure 3). This should in turn lower the acceptability of a missing NP sentence. The data, however, show that plausible missing NP sentences are relatively acceptable and V2 is not orphaned.

While we did not predict that missing NP sentences would feature such a set of thematic relations, these results turn out to be consistent with our thematic relations hypothesis, which attributes the missing VP/NP illusion to the parser establishing thematic relations between all arguments and predicates in the sentence. More specifically, in a missing NP sentence, all NPs (including the main clause subject) get a thematic role by virtue of being related to some verbal predicate. The subject of the sentence, appearing in a canonical subject position preceding V1, is interpreted as V1's experiencer or agent. The two post-verb NPs are interpreted as arguments of all three verbs, as ratings for the implausible conditions show. Conversely, every verbal predicate also gets at least one argument: V1 has an agent/experiencer in its subject and a theme argument in NP2; V3 has an agent in NP2 and a theme in NP1. While the experiment did not test whether V2 has an agent argument, our results indicate that it has a theme argument in NP2.

Finally, the illusion is not predicted by an account in which center-embedding illusions (or the lack thereof) are derived from language experience. Given the noun-final nature of Mandarin nominal expressions and the presence of noun clusters in the language, such an account predicts that Mandarin should lack the illusion, a point we return to in Section 7 (Experiment 4).

5 Experiment 2

Experiment 1 suggested that Mandarin has a missing NP illusion parallel to the missing VP illusion reported for English. However, as reviewers pointed out, this conclusion assumes that participants parsed missing NP sentences as being doubly center-embedded. Although we chose verbs and added adverbs to facilitate such a parse, one might still be concerned that some of these sentences were parsed as singly center-embedded. In that event, the absence of a third post-verbal NP would pose no structural problem, leading to higher acceptability.

As an illustration, in such a scenario, the missing NP sentence in (10) could receive the parse as bracketed in (16). Here, there is only one relative clause, with V2 *zebeiguo* "rebuked" as the verb of the clause. V2 takes a VP-like complement headed by V3, *shenli* "hear." NP2 *faguan* "judge" is interpreted as V2's relativized subject.

(16) Zongli zhengzai [$_{V1}$ jiejian] [$_{RC}$ ceng [$_{V2}$ zebeiguo] [$_{VP}$ gang [$_{V3}$ shenli]
prime.minister PROG meet previously rebuked recently hear
[$_{NP1}$ tanwu-an] bujiu]] de [$_{NP2}$ faguan].
corruption-case not.long.ago DE judge
Single center-embedding reading: '??The prime minister is meeting the judge who previously rebuked [the recent hearing of the corruption case not long ago].'

To the extent these parses are available, it implies that V2 verbs like "rebuked" can take VP-like complements in general, so structurally simpler sentences like (17) should also be acceptable. This seems unlikely. We had chosen verbs that we judged as incompatible with VP complements, especially when combined with a past-oriented time adverb. In our judgment, (17) is unnatural; the English translation is also rather awkward.

(17) Buzhang [$_{V2}$ zebeiguo] [$_{VP}$ gang [$_{V3}$ shenli] [$_{NP1}$ tanwu-an]].
minister rebuked recently hear corruption-case
Intended: '??The minister rebuked [the recent hearing of the corruption case].'

Nevertheless, to avoid prejudging verb complementation facts, we ran a follow-up acceptability judgment study to determine how acceptable sentences like (17) are. For thoroughness, we looked at all V2 and V1 verbs used in Experiment 1, just in case there also happened to be V1 verbs that can take VP-like complements.

Of course, it is difficult to know how to interpret acceptability ratings without baselines: how do we decide whether a verb like "rebuke" can take a VP-like complement? To address this question, we identified three verbs – *chengren* "admit (to)," *fouren* "deny," and *biaoshi* "say" – that can take VP-like complements containing past-oriented temporal adverbs (18). The goal was to aggregate the three verbs to form a baseline.

(18) Zongcai chengren [$_{VP}$ cengjing [$_{V2}$ weixie] [$_{NP2}$ zhuli]].
CEO admit previously threaten assistant
'The CEO admitted to having threatened the assistant previously.'

Given our choice of V1 and V2 verbs in Experiment 1, we expected them to be less acceptable on average than the baseline verbs. We also expected the V1 and V2 verbs to vary in acceptability, although we had no predictions about which verbs would be more acceptable.

5.1 Materials

We constructed six Subject–Main verb–Adverb–Embedded verb–NP sentences for each of the 42 unique V1 and V2 verbs used in Experiment 1. (In principle, there should have been 48 verbs, but six verbs occur in both V1 and V2 positions.) Within this set of six sentences, one (or two, for the six overlapping verbs) was constructed by reusing constituents from Experiment 1, so that the resulting sentence resembled the original grammatical DCE sentence. An example of this can be seen in (17), where the main verb is *zebeiguo* "rebuked": we used NP3 *buzhang* "minister" as the subject of this sentence, the adverb *gang* "recently," V3 *shenli* "hear" as the embedded verb, and NP1 *tanwu-an* "corruption case" as the NP. We added modal auxiliaries, aspect markers, or adverbs to the verbs of interest where necessary for felicity.

For the other sentences in each set, we replaced all lexical items other than the main verb and adverb. We made sure the replacement verbs and NPs were as plausible as possible, given the choice of main verb. In the context of (17), for instance, another sentence was constructed with *xuezhe* "scholar" as the subject, *tantao* "investigate" as the embedded verb, and *shehui wenti* "social issue" as the object. These are arguably as compatible with "*zebeiguo*" – a scholar might rebuke someone's investigation of social issues to the same extent as a government minister might rebuke someone's hearing of a corruption case. If these sentences are still unacceptable, we can then reasonably attribute that to structural factors.

We also created six sentences for each baseline verb, using the same Subject–Main verb–Adverb–Embedded verb–NP frame.

We wanted to keep the experiment short to help participants focus on the judgment task. To that end, we randomly sorted the 42 verbs into two non-overlapping sets of 21 verbs. We then added the three baseline verbs to each set, forming two sets of 24 verbs. Within each set of 24 verbs, the six sentences for each verb were then distributed to produce six lists, such that each verb appears once in the list. Put differently, there were a total of 12 lists (2 sets of 24 verbs x 6 lists per set), each containing 24 target sentences (one sentence for each of the 21 V1 and V2 verbs and three baseline verbs).

Finally, we added a set of 24 filler items to each list, so there were as many filler items as target and baseline sentences. The filler sentences were similar in length and structure to the other items, consisting of a subject, an adverb, an auxiliary or a second adverb, and a verb and an NP object. 21 of these sentences were grammatical, while the remaining three were not. On the assumption that the non-baseline target items were unacceptable, this distribution ensured that participants saw an equal number of acceptable and unacceptable sentences.

5.2 Participants

Participants were 36 native speakers of Mandarin on Prolific, recruited from the 60 participants who completed Experiment 1. The recruitment process was blind, in the sense that participants' demographic information and their responses from Experiment 1 were not taken into consideration. Each participant received US$1.60, based on our estimate of the time needed to complete the experiment and Prolific's recommended hourly rate.

5.3 Procedure

Recall that the V1 and V2 verbs were sorted into two sets and six lists of sentences were constructed for each set. We randomly assigned three participants to each list, so that a total of 18 participants rated sentences for each set of verbs. In other words, through this arrangement, we collected 18 responses for each V1 and V2 verb.

As in Experiment 1, participants rated the sentences on Ibex, with the same 7-point Likert scale. Before starting the experiment, they saw seven practice sentences. For the first four practice sentences, participants saw brief comments about each sentence and what kind of ratings to assign. Instead of a missing NP sentence, the fourth practice sentence here was a sentence involving an apparent subcategorization violation (an intransitive verb occurring with a post-verb NP).

5.4 Data analysis

After data collection, we noticed errors in three of the target sentences (one featuring the verb *dao* "visit, go to" and two featuring *jiaru* "join"); these sentences have been excluded from analysis. To ensure that participants had taken the time to read a target sentence before rating it, we also excluded responses provided in less than one second after the sentence was presented. Altogether, 1.2% of responses were excluded.

We analyzed the results with a cumulative link mixed effects model for each set of verbs. There were two conditions (baseline verbs vs. V1 and V2 verbs), by-subject and by-verb random intercepts, and a by-subject random slope in each model.

5.5 Results

Table 3 presents a summary of acceptability ratings. The sentences for the three baseline verbs received uniformly very high ratings, while the sentences for the V1 and V2 verbs received much lower ratings. Statistical analysis confirmed that the difference is significant ($p < 0.001$, Table 4). There was also some variation within the class of V1 and V2 verbs, but there was no single verb whose ratings were comparable to the baseline verbs.

Table 3: Mean acceptability ratings for Experiment 2, aggregated across both sets.
(1 = completely unacceptable, 7 = completely acceptable)

Condition	Mean acceptability rating (SE)
Baseline verbs	6.62 (0.07)
All V1 and V2 verbs	2.43 (0.06)
	Maximum: 3.33 (0.30), for *diaocha* "investigate"
	Minimum: 1.72 (0.23), for *genzong* "trail, follow"

Table 4: Mixed effects regression results.
(baseline: baseline verbs *chengren* "admit," *fouren* "deny" and *biaoshi* "say")

Condition	Estimate	SE	z value	p value
First set of V1 and V2 verbs	-10.36	1.53	-6.78	<0.001
Second set of V1 and V2 verbs	-7.54	1.02	-7.41	<0.001

5.6 Discussion

The results provide clear evidence that the V1 and V2 verbs in Experiment 1 do not allow VP complements containing a past-oriented time adverb. It is therefore unlikely that the missing NP sentences in Experiment 1 were relatively acceptable because some of those sentences had a grammatical single center-embedding parse that is contingent on V1 or V2 verbs taking such a VP complement.

To sum up: the results of Experiments 1 and 2 jointly indicate that Mandarin Chinese exhibits a center-embedding illusion analogous to the missing VP illusion reported for English. The implausibility manipulations in Experiment 1 further reveal an interesting difference: in a missing NP sentence, the two NPs that appear after the verbs get interpreted as objects and subjects of the preceding verbs in a way that lacks a parallel in English missing VP sentences. These interpretations are also not predicted by existing accounts of the center-embedding illusion.

However, these remarks presume that our description of how missing VP sentences are interpreted is accurate. So that we have a stronger basis for comparing Mandarin with English, we ran a replication of Gibson & Thomas 1999. Our goal here was not to test their claims about structural forgetting. Rather, we were interested in using their manipulations to run an experiment parallel to Experiment 1: verifying whether English exhibits the missing VP illusion, and if so, which NPs serve as the arguments of which verb in a missing VP sentence.

6 Experiment 3

6.1 Materials

Experimental materials were based on the 12 items listed on pp. 247-248 of Gibson & Thomas 1999. These were used to generate sentences for the four conditions in (3), repeated below as (19). We follow Gibson & Thomas in labeling these conditions as grammatical DCE, missing VP1, missing VP2, and missing VP3 sentences.

(19) Experiment items from Gibson & Thomas 1999

a. Grammatical DCE
[$_{NP1}$ The novel] that [$_{NP2}$ the horror author] who [$_{NP3}$ the publishing company] [$_{VP1}$ had recently fired] [$_{VP2}$ had typed quickly] [$_{VP3}$ was banned by the local library].

b. Missing VP1
[$_{NP1}$ The novel] that [$_{NP2}$ the horror author] who [$_{NP3}$ the publishing company] ---- [$_{VP2}$ had typed quickly] [$_{VP3}$ was banned by the local library].

c. Missing VP2
[$_{NP1}$ The novel] that [$_{NP2}$ the horror author] who [$_{NP3}$ the publishing company] [$_{VP1}$ had recently fired] ---- [$_{VP3}$ was banned by the local library].

d. Missing VP3
[$_{NP1}$ The novel] that [$_{NP2}$ the horror author] who [$_{NP3}$ the publishing company] [$_{VP1}$ had recently fired] [$_{VP2}$ had typed quickly] ----.

As noted previously, the three missing VP conditions feature plausibility manipulations to confirm claims that speakers treat the third NP as the first VP's subject, while treating the first NP

(but not the second NP) as the subject of the second VP. If so, one expects a plausibility violation in the missing VP1 condition (19b): the first VP *had typed quickly* would take the third NP *the publishing company* as an agent. A similar violation would occur in the missing VP3 condition (19d), where the second VP, *had typed quickly*, takes the first NP *the novel* as its agent.

The 12 test sentences were distributed across four lists in a Latin Square design, so that in each list were three sentences per condition, and no two sentences were variants of each other. Following Gibson & Thomas, test sentences were mixed with 48 filler sentences of comparable length and complexity, featuring various kinds of adjunct, relative, and complement clauses. 30 of these sentences were well-formed and acceptable and 18 were ill-formed and unacceptable. This ratio was chosen so that participants would see as many acceptable sentences as unacceptable sentences, on the assumption that all four experimental conditions are relatively unacceptable.

We also prepared 7 practice items. Three of these items were grammatical and comprehensible while the other four items were not. One of the four ungrammatical practice sentences was the following non-DCE sentence missing a VP, used by Gibson & Thomas as a practice sentence: *The form was stamped by the bureaucrat who worked at the ministry where everyone who had walked strangely.*

6.2 Participants

Participants were 32 US-based workers on Amazon Mechanical Turk. All 32 participants were self-identified native speakers of American English and had passed a native speaker proficiency test, which tested for knowledge of relatively subtle grammatical rules and constraints of American English. Participants received US$2.50.

6.3 Procedure

The procedure was largely the same as that of Experiment 1. We departed from Gibson & Thomas in our choice of scale. Gibson & Thomas used a 1 to 5 scale, where "5" indicated "hard to understand" and "1" "easy to understand." To be consistent with Experiment 1, we used a 7-point scale and reversed it, so "1" was used for unacceptable sentences and "7" for acceptable ones.

For familiarization purposes, participants first judged the practice sentences. For the first three practice sentences, participants saw brief comments about each sentence and what kind of ratings to assign, e.g. a sentence should have received a low rating because it described an implausible scenario. The fourth practice sentence was Gibson & Thomas's non-DCE sentence that lacked a VP. Following Gibson & Thomas, the comment for the missing VP sentence indicated that they should give similar sentences a low rating, although it did not explain why.

6.4 Data analysis

To ensure that participants took the time to read a target sentence before judging it for acceptability, we excluded responses that were provided in 3 seconds or less upon the presentation of the sentence; doing so eliminated about 6% of responses.

As was the case for Experiment 1, ratings were analyzed with a cumulative link mixed effects model, with conditions as a fixed effect and random intercepts and slopes for both participant and items. Ratings for the missing VP2 condition were used as a baseline.

6.5 Results

Table 5 shows that in absolute terms, all four conditions received low to medium acceptability ratings. Of the three missing VP conditions, the missing VP2 condition was rated the most acceptable; this outcome was also observed by Gibson & Thomas and Frank & Ernst (2019).

Table 5: Mean acceptability ratings for Experiment 3.
(1 = completely unacceptable, 7 = completely acceptable)

Condition	Mean acceptability rating (standard error)
Grammatical DCE (19a)	3.10 (0.17)
Missing VP1 (19b)	2.47 (0.16)
Missing VP2 (19c)	3.29 (0.17)
Missing VP3 (19d)	2.41 (0.15)

Statistically, the missing VP2 condition was significantly more acceptable than the missing VP1 condition ($p = 0.001$) and the missing VP3 condition ($p < 0.001$). Although the missing VP2 condition also received numerically higher ratings than the grammatical DCE condition, the difference is not significant ($p = 0.357$).

Table 6: Mixed effects regression results.
(baseline: missing VP2 condition)

Condition	Estimate	SE	z value	p-value
Grammatical DCE	-0.28	0.30	-0.92	0.357
Missing VP1	-1.44	0.44	-3.25	0.001
Missing VP3	-1.55	0.39	-3.95	<0.001

6.6 Discussion

Our results replicated Gibson & Thomas's key findings. The fact that ungrammatical missing VP2 sentences were not less acceptable than grammatical DCE sentences indicates a grammaticality illusion. We also confirmed their observations about missing VP sentences: the first VP has a subject in the third NP, while the second VP has a subject in the first NP but not the second NP. In terms of thematic relations, the latter result means that the second VP is related to the first NP, but not the second. This result contrasts with the finding for Mandarin missing NP sentences in Experiment 1, where the second post-verb NP (the equivalent of the second VP in English) is thematically related to both the first and second verbs (the equivalent of the first and second NPs in English).

To sum up, Experiments 1 and 3 show that ungrammatical DCE sentences with missing NPs as VPs in Mandarin and English exhibit a grammaticality illusion, even though these ungrammatical sentences are assigned very different thematic relations.

These findings have implications for cross-linguistic accounts of center-embedding illusions. Ideally, an account must be fine-grained enough to explain the differences in thematic relations between Mandarin and English. However, it also needs to be able to capture a higher-level generalization, namely, that Mandarin and English speakers experience a grammaticality illusion, while German and Dutch speakers are less likely to do so. In the next section, we address

the second issue about the cross-linguistic distribution of the illusion. In particular, we evaluate an approach that attributes the distribution to cross-linguistic differences in language experience (Engelmann & Vasishth 2009; Vasishth et al. 2010; Frank et al. 2016; Futrell & Levy 2017; Frank & Ernst 2019; Futrell et al. 2020).

7 Experiment 4

The language experience hypothesis was proposed to explain why not all languages exhibit a center-embedding illusion to the same degree. Under this hypothesis, the verb-final syntax of German and Dutch means that long dependencies between subjects and verbs and verb sequences occur relatively frequently. Exposure to these linguistic structures is argued to help German and Dutch speakers process DCE sentences more easily.

We argued that Mandarin presents a problem for this hypothesis because its noun phrases are consistently head-final, which parallels the verb-final syntax of German and Dutch. For this reason, long dependencies between verbs and the head noun of objects are relatively frequent, as are sentences containing sequences of nouns. The language experience hypothesis therefore predicts that Mandarin should pattern like German and Dutch.[7]

In this section, we make this argument more explicitly, using Futrell et al.'s (2020) noisy surprisal model of the illusion (also Futrell & Levy 2017). This model was designed to capture differences between German and English in the center-embedding illusion, based on the language experience hypothesis. We adapt the model for Mandarin Chinese and show that it incorrectly predicts that Mandarin should pattern like German and Dutch.

7.1 The model

We chose Futrell et al.'s model over the simple recurrent neural network (RNN) models used by Christiansen, Vasishth, Frank, and colleagues for three reasons. First, as mentioned, Futrell et al.'s model is intended as an implementation of the language experience hypothesis: it explicitly derives the presence and absence of the illusion in English and German as a consequence of the fact that German relative clauses are always verb-final, while English relative clauses are not. Second, the code is publicly available, which lets us reproduce their analysis faithfully. Third, as they argue, RNN models can be thought of as a "special case" of a lossy-context surprisal model (Futrell et al. 2020:17--18).

While Futrell et al.'s model was presented as a model of reading times for DCE sentences (Vasishth et al. 2010), we think that it is reasonable to use it to model acceptability judgments.

[7] A reviewer asked if this particular prediction might simply follow from the type of relative clauses in DCE sentences, rather than from language experience: English DCE sentences feature object relatives; German and Dutch DCE sentences feature either object or subject relatives. Not unlike German and Dutch, Mandarin DCE sentences feature subject relatives. To the extent that subject relatives have a processing advantage, one might expect German, Dutch, and Mandarin DCE sentences to be easier to process and more acceptable. (Note that there is a debate on whether subject relatives have a processing advantage, especially for Chinese. See Hsiao & Gibson 2003; Lin & Bever 2006, 2011; Vasishth et al. 2013; Jäger et al. 2015.)

We are less certain about this line of reasoning. Even if there is a genuine subject relative processing advantage, it is unclear whether this processing advantage would persist under double center-embedding. It is not implausible that the structural complexity of double center-embedding would neutralize the processing advantage.

Frank & Ernst (2019) note that Dutch and German have similar verb-final syntax and DCE sentences in both languages have the same reading time profiles. They further point out that reading time differences between grammatical DCE and missing VP sentences in Dutch and English correlate with an acceptability difference.

Futrell et al.'s model uses a simple probabilistic grammar, simulating English and German as a set of intransitive sentences in which the subject is sometimes separated from the verb by modifiers. The focus on intransitive sentences is presumably driven by the goal to model the parsing of a prototypical DCE sentence, in which a subject NP is separated from the verb by a relative clause containing another relative clause.

Table 7: Toy grammar rules.

English / German	Probability	Mandarin	Probability
S → NP V	1	S → N V NP	1
NP → N	0.5	NP → N	0.5
NP → N RC	0.25	NP → RC N	0.25
NP → N PP	0.25	NP → N ConjP	0.25
PP → P NP	1	ConjP → Conj NP	1
RC → C NP V	0.2 (1 for German)	RC → NP V C (assuming C = *de*)	0.2
RC → C V NP	0.8 (0 for German)	RC → V NP C	0.8

Note: Mandarin PP modifiers cannot follow nouns. To capture this fact while ensuring that our Mandarin rules parallel Futrell et al.'s rules, we replace the NP → N PP rule with another rule that conjoins a noun with an NP. This change is effectively cosmetic, with no bearing on the strings generated.

In this grammar (Table 7), a noun is modified by a relative clause 25% of the time. For English, it is stipulated that 20% of relative clauses are verb-final, reflecting the fact that verb-final object relative clauses are infrequent in naturalistic contexts. For German, all relative clauses are verb-final. With these parameters, the grammar generates a set of sequences of syntactic categories (corresponding to sentences) and their probabilities. Noise is then applied to these sequences, so that each syntactic category is deleted with a probability of 20%, producing sequences of varying grammaticality. It is this noisier set of sequences that the model is exposed to.

We extend the model with a simple probabilistic grammar of Mandarin. Since the prototypical Mandarin DCE sentence features a transitive verb separated from its object by a doubly embedded relative clause, we have this grammar generate transitive sentences instead. Following Futrell et al., we assume that 20% of relative clauses are object relative clauses, and that the remainder are subject relative clauses.

One might wonder whether these parameter values are realistic. With these values, the grammars for all three languages assign low, but non-zero, probabilities to grammatical DCE sentences before noise is applied, even though DCE sentences are likely to be effectively non-existent in naturalistic contexts. However, this model is arguably not intended to show how actual differences in language statistics can derive differences in how DCE sentences are parsed. Rather, it is a proof of concept, demonstrating how the parsing of DCE sentences can be conditioned by general word order differences, such as whether relative clauses are uniformly verb-final. These parameter values serve as simplifying assumptions necessary for the demonstration of this point.

7.2 Model results

We first ran the model for English and German/Dutch to replicate Futrell et al.'s results. Doing so also allows us to confirm that their choice of parameter values would be appropriate for Mandarin, for the sake of fairness. Having done that, we ran the model for Mandarin.

The English and German/Dutch models were presented with the DCE sequence NCNCNVV (N = noun, V = verb, C = relative pronoun). The model then calculated the probability that this sequence is followed by a V or by an end-of-sentence symbol. To the extent that the English model generates the illusion, the surprisal of observing the end-of-sentence symbol should be lower than the surprisal for a third verb, even though the English grammar generates DCE sentences. We predict the opposite contrast for the German/Dutch model. These predictions are borne out (Table 8).

The Mandarin model was presented with the DCE sequence NVVVNCN (N = noun, V = verb, C = *de*), and it calculated the probability that this sequence would be followed by a C (*de*) or by the end-of-sentence symbol. As mentioned, the language experience hypothesis should predict the absence of an illusion: the surprisal of observing C (*de*), i.e. the grammatical continuation, should be lower. The model bears out this prediction.

Table 8: Noisy surprisal model results (in bits) for English, German/Dutch, and Mandarin.

Language	Surprisal for grammatical continuation	Surprisal for ungrammatical continuation	Difference (Grammatical – Ungrammatical)
English	1.905	1.068	0.838
German/Dutch	0.935	1.964	-1.029
Mandarin	1.822	2.583	-0.762

We also ran a post-hoc analysis for Mandarin, varying the conditional probability of subject relative clauses. Doing so let us assess the alternative hypothesis that the observed illusion in Mandarin is just because subject relative clauses are less common; Hsiao & Gibson (2003) estimate the rate of subject relative clauses to be about 57.5% (although Vasishth et al. (2013) estimate it to be higher, around 73%). As Table 9 shows, the model predicts that Mandarin only starts exhibiting the illusion when the probability of subject relative clauses falls below 20%, which is unrealistically low. We therefore discount this alternative hypothesis.

Table 9: Mandarin surprisal difference and probability of subject relative clauses.

P(subject relative clauses | relative clause)	80% (see Table 8)	60%	40%	20%
Surprisal difference	-0.762	-0.616	-0.395	0.027

An anonymous reviewer raised the possibility that the results might be an artifact of the deletion rate. To address this issue, we reran the model with different deletion rates. Results of this post-hoc analysis are presented in Table 10. In general, when the deletion rate is low (e.g. below 0.3), the model predicts that Mandarin lacks the illusion, patterning like German / Dutch: grammatical continuations elicit a lower surprisal than ungrammatical continuations. When the deletion rate is higher, the model predicts that Mandarin starts exhibiting the illusion, like English.

Table 10: Surprisal difference and deletion probability.

Deletion rate	English	German/Dutch	Mandarin
0.1	0.102	-1.888	-2.771
0.2 (default)	0.838	-1.029	-0.762
0.3	1.006	-0.640	0.084
0.4	0.934	-0.437	0.558
0.6	0.495	-0.257	1.116
0.8	-0.039	-0.271	1.360

One possible takeaway is that the deletion rate should have been higher. For instance, contrary to our current assumptions (and Futrell et al.'s), perhaps the deletion rate varies with processing load, so that the deletion rate is higher for DCE sentences than for simpler sentences. On the surface, this is not an implausible scenario. However, it calls for a richer theory of deletion and representations beyond what is now available. For instance, one would need to specify when in incremental processing and by how much the deletion rate starts to rise (cf. Gibson & Thomas 1999). More generally, there are also potential consequences regarding how missing VP/NP sentences are interpreted, since this operation ostensibly deletes words from a linguistic representation. As far as we can tell, though, these issues have not been spelled out, so it is unclear how we can explain existing findings. For these reasons, we will not adopt this alternative perspective here, but flag it for future research.

8 General discussion

To recap, the present study evaluated existing accounts of center-embedding illusions with novel Mandarin Chinese data. Experiments 1 and 3 confirmed that Mandarin and English both exhibit a center-embedding illusion. However, both experiments also show that the nouns and verbs of these sentences are related to each other in strikingly different ways. As shown in Figure 5, in English, the final VP is linked to only the first NP. In Mandarin, however, the final NP appears to be linked not only to the first verb, but also to the second and third verbs.

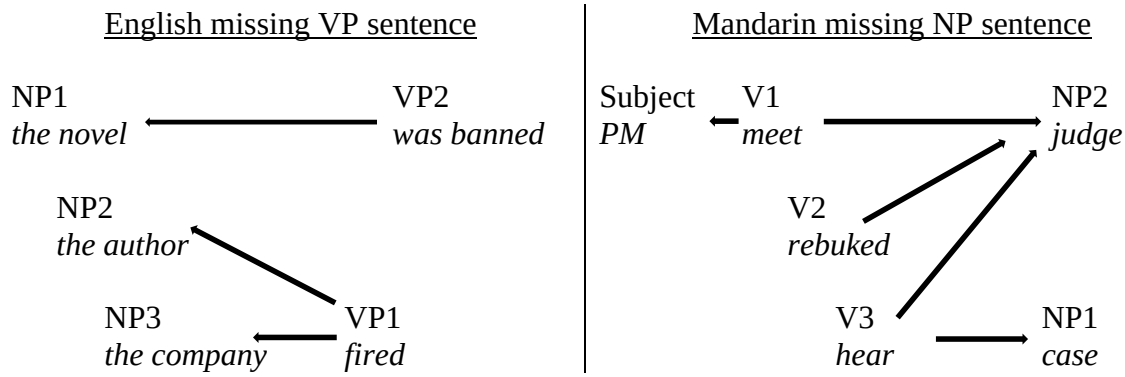


Figure 5: Thematic relations in missing VP/NP sentences.

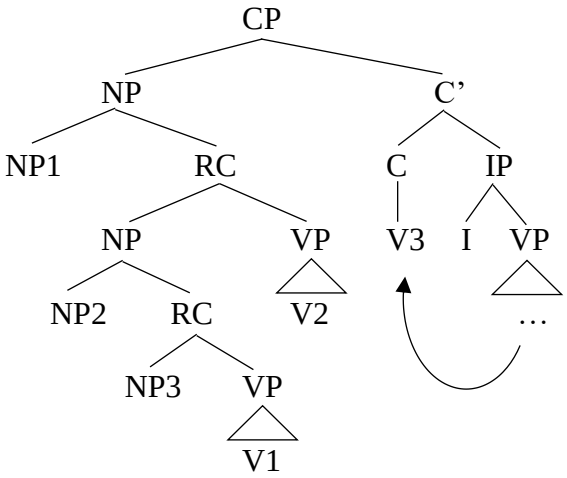
Despite this difference, there is a more abstract way in which both types of sentences are the same: there are thematic relations connecting all arguments and predicates; no NP or verb is "orphaned" thematically. The Mandarin and English results are thus consistent with the thematic

relations hypothesis, which posits that the presence of such relations produces the sense of completeness that characterizes the missing VP/NP illusion.

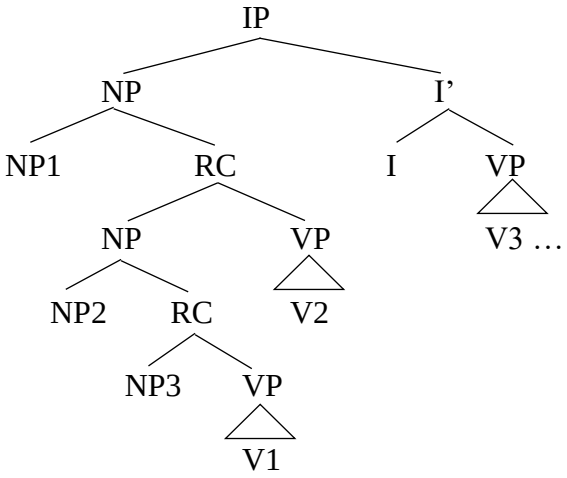
Mandarin also has implications for theories about the cross-linguistic distribution of the illusion. In Experiment 4, we argued that the Mandarin illusion poses a challenge for accounts that explain cross-linguistic variation by appealing to the relative frequency of verb-final (or noun-final) structures in a language (Vasishth et al. 2010; Frank et al. 2016; Futrell et al. 2020). Here, we consider an alternative account, namely, the interference account of Bader (2016), exploring how it can be adapted to accommodate the Mandarin data.

To recap, Bader assumes that the parser builds different syntactic representations for string-equivalent German and English DCE sentences. The verb of a German (and Dutch) main clause appears in what is traditionally known as a verb-second position. In the formalism that Bader adopts, this verb (V3) is found in C, having moved from a VP position, as shown in (20a). The other two verbs, on the other hand, remain in VPs in their respective clauses. In contrast, English verbs all remain in their respective VPs (20b).

(20) a. German (adapted from Bader 2016, ex. 22)



b. English (adapted from Bader 2016, ex. 17b)



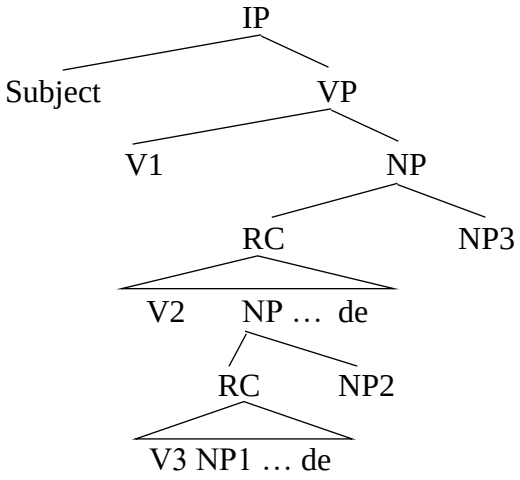
According to Bader, this syntactic difference has an impact on how the parser processes the second verb. In English, both the V2 and V3 slots are inside VPs, so the parser is likely to

confuse them and mis-attach the verb to the V3 slot, producing the classic missing VP effect. Mis-attachment to the V3 slot is even likelier if this slot enjoys a primacy advantage over the V2 slot, as Häussler & Bader (2015) suggest. In contrast, the risk of mis-attachment is lower in German. The two slots are structurally distinct – the V2 slot is inside a verb-final VP, while the V3 slot is inside C. As a result, the parser is less likely to confuse them.

The above description of (mis)attachment can be implemented in terms of memory retrieval and interference (Lewis & Vasishth 2005; McElree et al. 2003; McElree 2000, etc.), with the assumption that retrieval cues can refer to the structural features of the attachment site, such as the node that immediately dominates the site (i.e. VP or C). Upon observing the second verb, the parser needs to retrieve the correct verb attachment site from memory. By hypothesis, it sets one of the retrieval cues to be [VP]. In English, since there are two VPs with empty V slots, the parser might incorrectly retrieve the main clause VP instead. In contrast, in German, there is only one VP with an empty V slot, so the parser is less likely to make a mistake.

The relevant syntactic facts for Mandarin resemble English more strongly than German: Mandarin does not syntactically distinguish between objects in main clauses and subordinate clauses (21). We suggest that this might explain why the illusion is present in Mandarin and English. Ideally, when processing a Mandarin DCE sentence, the parser should attach the second post-verb noun to the NP2 slot and not the NP3 slot in (20). However, like V2 and V3 in English, both slots are structurally similar: they are both NP slots dominated by another NP node. The Mandarin parser is thus likely to confuse them and mis-attach the noun into the NP3 slot instead.

(21) Mandarin



However, as discussed in Section 3.3, this analysis only predicts that V1 and V3 have arguments; V2 might still be "orphaned" thematically. This is inconsistent with the results of Experiment 1, which show that the second noun gets interpreted as a theme argument (an object) of V2, in addition to being the agent of V3 and the theme of V1. A purely syntactic approach also cannot capture these thematic relations. There is no well-formed double center-embedding syntactic representation where the second noun receives such an interpretation.

We suggest supplementing the analysis with the assumption that the parser tracks thematic relations throughout sentence processing and repairs them when necessary. In English, repair is likely to be unnecessary, since mis-attachment of the second verb does not result in any of the VPs or NPs becoming thematically orphaned. In contrast, in Mandarin, mis-attachment causes V2 to

be thematically orphaned. To repair this problem, the parser proceeds to link V2 to NP2 (Figure 6). Successful repair therefore ensures that there are thematic relations between all arguments and predicates, producing the same illusion of completeness.

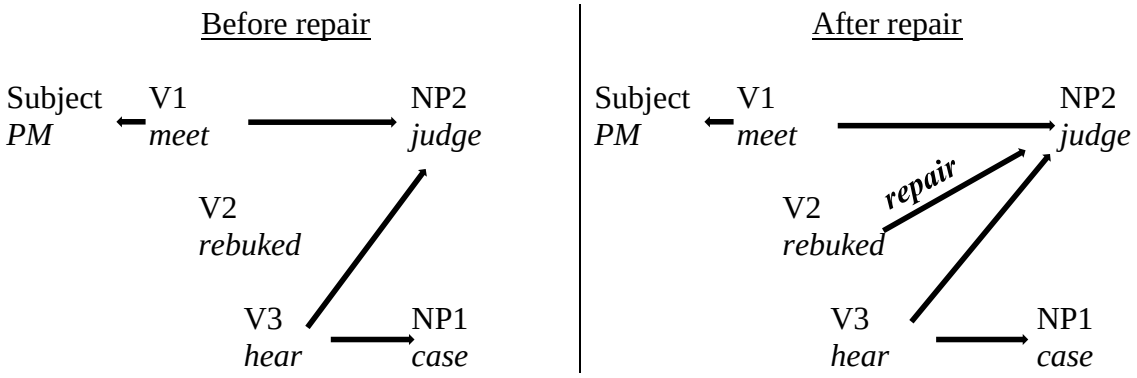


Figure 6: Proposed repair of thematic relations in Mandarin.

Since there is no way to derive these thematic relations syntactically, it is unlikely that the parser is repairing a syntactic representation like the one in (21). These thematic relations therefore implicate the presence of a second representation for missing NP sentences, presumably a non-syntactic one. This is consistent with a suggestion of ours in Section 2.2. There, we argued that although the parser might build a fully articulated syntactic representation of a DCE sentence, for memory load reasons, it might not be able to reliably use such a representation to determine thematic relations. Instead, the parser might rely on a shallower representation that directly tracks the arguments and predicates in the sentence. Such a representation could also make repairs easier to implement.

In principle, this repair scenario predicts differences in how quickly speakers might detect implausibility in the various missing NP conditions in Experiment 1, where verbs were chosen so that NP1 and NP2 would be inappropriate arguments. More specifically, the repair scenario predicts that speakers might take longer to notice that NP2 is an inappropriate argument for V2, since the relation between NP2 and V2 is built late. This prediction, being a claim about timing, is difficult to verify with the offline acceptability judgment ratings collected in Experiment 1. We leave it to future work to test this prediction using methods that yield better temporal resolution, such a self-paced reading paradigm.

9 Conclusion

We demonstrated that Mandarin speakers experience a missing NP illusion in processing sentences with double center-embedding, analogous to the better-studied missing VP illusion in languages like English (Gibson & Thomas, 1999; see Gimenes et al. 2009 for French; also see Häussler & Bader 2015; Bader 2016 for relevant German data). We noted that the interpretations of missing NP sentences do not follow easily from structural forgetting accounts like Gibson & Thomas's and interference accounts like Häussler and Bader's. Mandarin also presents a challenge for language experience hypotheses, which are intended to account for the cross-linguistic variation in this illusions (Vasishth et al. 2010; Frank et al. 2016; Futrell & Levy 2017; Frank & Ernst 2019; Futrell et al. 2020). These hypotheses predict that Mandarin should lack the illusion.

Our experiment results suggested that the exceptional acceptability of missing VP and NP sentences is likely related to the fact that speakers were able to build thematic relations between all arguments and predicates denoted by the NPs and VPs in these sentences. In light of our modeling results showing that the language experience hypothesis predicts the absence of the missing NP illusion, we discussed adapting an alternative interference hypothesis of Bader (2016). We suggested that the interference approach, supplemented with a repair mechanism, can better explain why such the illusion surfaces in languages like English and Mandarin, but not in languages like German and Dutch.

Abbreviations

Abbreviations in glosses follow the Leipzig glossing rules, except for: EXP: experiential aspect (-*guo*); DE: the particle *de* for modifying nouns.

Supplementary Files

1. Materials for Experiments 1 and 2
2. Python script for Experiment 4

Ethics and consent

Experiments were approved by [anonymized for review] (reference number [anonymized]). Informed consent was obtained from all experiment participants.

Funding information

[anonymized for review]

Acknowledgements

[anonymized for review]

Competing interests

The authors have no competing interests to declare.

References

Aoshima, Sachiko, Colin Phillips & Amy Weinberg. 2004. Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language* 51(1). 23–54. https://doi.org/10.1016/j.jml.2004.03.001

Aoshima, Sachiko, Masaya Yoshida & Colin Phillips. 2009. Incremental processing of coreference and binding in Japanese. *Syntax* 12(2). 93–134. https://doi.org/10.1111/j.1467-9612.2009.00123.x

Bader, Markus. 2016. Complex center embedding in German – The effect of sentence position. In Sam Featherston & Yannick Versley (eds.), *Quantitative approaches to grammar and grammatical change: perspectives from Germanic*, 9-32. Berlin: De Gruyter.

Boland, Julie E., Tanenhaus, Michael K., Garnsey, Susan M., & Carlson, Greg N. 1995. Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language*. 34(6). 774–806. https://doi.org/10.1006/jmla.1995.1034

Brasoveanu, Adrian & Jakub Dotlačil. 2020. *Computational Cognitive Modeling and Linguistic Theory*. Cham: Springer.

Chomsky, Noam, & George A. Miller. 1963. Introduction to the formal analysis of natural languages. In R. Duncan Luce, Robert R. Bush and Eugene Galanter (eds.), *Handbook of mathematical psychology*, vol. 2, 269-321. New York: Wiley.

Christiansen, Morton H. & Nick Chater. 1999. Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23(2). 157–205. https://doi.org/10.1207/s15516709cog2302_2

Christiansen, Morten H. & MacDonald, Maryellen C. 2009. A usage-based approach to recursion in sentence processing. *Language Learning* 59. 126–161. https://doi.org/10.1111/j.1467-9922.2009.00538.x

Christensen, Rune H. B. 2019. ordinal—regression models for ordinal data. R package version 2019. 4-25. http://www.cran.r-project.org/package=ordinal/.

Cowper, Elizabeth A. 1976. *Constraints on sentence complexity: A model for syntactic processing*. Brown University doctoral dissertation.

Engelmann, Felix & Shravan Vasishth. 2009. Processing grammatical and ungrammatical center embeddings in English and German: A computational model. In Andrew Howes, David Peebles, & Richard Cooper (eds.), *Proceedings of 9th International Conference on Cognitive Modeling*, 240–245.

Fodor, Janet D., Stefani Nickels & Esther Schott. 2017. Center-embedded sentences: What's pronounceable is comprehensible. In Roberto G. de Almeida and Lila R. Gleitman (eds.), *On concepts, modules, and language: Cognitive science at its core*, 139-168. New York: Oxford University Press.

Frank, Stefan L. & Patty Ernst. 2019. Judgements about double-embedded relative clauses differ between languages. *Psychological Research* 83. 1581-1593. https://doi.org/10.1007/s00426-018-1014-7.

Frank, Stefan L. Thijs Trompenaars & Shravan Vasishth. 2016. Cross-linguistic differences in processing double-embedded relative clauses: Working-memory constraints or language statistics? *Cognitive Science* 40(3). 554-578. https://doi.org/10.1111/cogs.12247

Frazier, Lyn. 1985. Syntactic complexity. In David R. Dowty, Laurie Karttunen and Arnold Zwicky (eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives*, 129-189. New York: Cambridge University Press.

Futrell, Richard & Roger Levy. 2017. Noisy-context surprisal as a human sentence processing cost model. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 688–698.

Futrell, Richard, Edward Gibson & Roger Levy. 2020. Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science* 44(3). 1-54. https://doi.org/10.1111/cogs.12814

Gibson, Edward. 1998. Syntactic complexity: Locality of syntactic dependencies. *Cognition* 68(1). 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gibson, Edward & James Thomas. 1999. Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes* 14(3). 225-248. https://doi.org/10.1080/016909699386293

Gimenes, Manuel, François Rigalleau & Daniel Gaonac'h. 2009. When a missing verb makes a French sentence more acceptable. *Language and Cognitive Processes* 24(3). 440-449. https://doi.org/10.1080/01690960802193670

Häussler, Jana & Markus Bader. 2015. An interference account of the missing VP effect. *Frontiers in Psychology* 6:766. https://doi.org/10.3389/fpsyg.2015.00766

Hsiao, Franny & Edward Gibson. 2003. Processing relative clauses in Chinese. *Cognition* 90(1). 3–27. https://doi.org/10.1016/S0010-0277(03)00124-0

Huang, C.-T. James. 1984. On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15. 531–574.

Jäger, Lena, Zhong Chen, Qiang Li, Chien-Jer Charles Lin & Shravan Vasishth. 2015. The subject-relative advantage in Chinese: Evidence for expectation-based processing. *Journal of Memory and Language* 79-80. 97–120. https://doi.org/10.1016/j.jml.2014.10.005

Kazanina, Nina, Ellen Lau, M. Lieberman, Masaya Yoshida & Colin Phillips. 2007. The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language* 56(3). 384–409. https://doi.org/10.1016/j.jml.2006.09.003

Karlsson, Fred. 2007. Constraints on multiple center-embedding of clauses. *Journal of Linguistics* 43(2). 365-392.

Lewis, Richard L. & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29(3). 375–419. https://doi.org/10.1207/s15516709cog0000_25

Li, Y.-H. Audrey, 2014. Born empty. *Lingua* 151. 43–68.

Lin, C.-J. Charles & Thomas G. Bever. 2006. Subject preference in the processing of relative clauses in Chinese. In Donald Baumer, David Montero, and Michael Scanlon (eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics*, 254–260. Somerville, MA: Cascadilla Press.

Lin, C.-J. Charles & Thomas Bever. 2011. Garden path and the comprehension of head-final relative clauses. In Hiroko Yamashita, Yuki Hirose, and Jerome L. Packard (eds.), *Processing and Producing Head-final Structures*, 277–297. Dordrecht: Springer.

McElree, Brian. 2000. Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research* 29(2). 155–200. https://doi.org/10.1023/A:1005184709695

McElree, Brian, Stephani Foraker & Lisbeth Dyer. 2003. Memory structures that subserve sentence comprehension. *Journal of Memory and Language* 48(1). 67–91. https://doi.org/10.1016/S0749-596X(02)00515-6

Miller, George A. & Stephen Isard. 1964. Free recall of self-embedded English sentences. *Information and Control* 7, 292–303.

Phillips, Colin. 2006. The real–time status of island phenomena. *Language* 82(4). 795–823.

Pritchett, Bradley L. 1988. Garden path phenomena and the grammatical basis of language processing. *Language* 64(3). 539–576.

Pulman, S. G. 1986. Grammars, parsers, and memory limitations. *Language and Cognitive Processes* 1(3). 197-225. https://doi.org/10.1080/01690968608407061

R Development Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Resnik, Philip. 1992. Left-corner parsing and psychological plausibility. In *Proceedings of the 14th Conference on Computational Linguistics, Volume 1*, 191-197. Association for Computational Linguistics.

Stowe, Laurie A. 1986. Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1(3). 227–245. https://doi.org/10.1080/01690968608407062

Sturt, Patrick. 2003. The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language* 48(3). 542–562. https://doi.org/10.1016/S0749-596X(02)00536-3

Traxler, Matthew J. & Martin J. Pickering. 1996. Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35(3). 454–475. https://doi.org/10.1006/jmla.1996.0025

Vasishth, Shravan, Katja Suckow, Richard L. Lewis & Sabine Kern. 2010. Short-term forgetting in sentence comprehension: crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes* 25(4). 533–567. https://doi.org/10.1080/01690960903310587

Vasishth, Shravan, Zhong Chen, Qiang Li & Gueilan Guo. 2013. Processing Chinese relative clauses: Evidence for the subject-relative advantage. *PLoS ONE* 8(10). E77006. https://doi.org/10.1371/journal.pone.0077006

Wagers, Matthew & Colin Phillips. 2009. Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics* 45. 395–433. https://doi.org/10.1017/S0022226709005726

Xiang, Ming, Brian Dillon & Colin Phillips. 2009. Illusory licensing effects across dependency types: ERP evidence. *Brain and Language* 108(1). 40–55. https://doi.org/10.1016/j.bandl.2008.10.002

Xue, Nianwen & Martha Palmer. 2008. Adding semantic roles to the Chinese Treebank. *Natural Language Engineering*, 15(1). 143–172.