When missing NPs make double center-embedding sentences acceptable

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

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 rather 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) robustly 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 proposals have in turn attributed this variation to word order statistics in these languages, suggesting that linguistic experience can influence the processing of DCE sentences and the magnitude of the illusion (Vasishth et al. 2010; Frank et al. 2016; Futrell & Levy 2017; Frank & Ernst 2019; Futrell et al. 2020).

In this paper, we are 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 illusion 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 &

---

[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).

Levy 2017), extending it with the goal of modeling the differences reported between German, Dutch, English and Mandarin Chinese.

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 6. We discuss implications of our results in Section 7, before concluding in Section 8.

2 The Missing VP illusion

In this section, we consider what might cause missing VP sentences in languages like English to be perceived as relatively acceptable compared to their grammatical counterparts. For scope reasons, we will restrict ourselves, in Section 2.1, to reviewing recent accounts that attribute the illusion to memory-related factors, which have been adapted to explain cross-linguistic variation.[2] We distinguish between three classes of accounts, referring to them here as the 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 break down this question into two smaller questions. The first question is about representations: what is the relevant representation in which a VP is omitted and over which acceptability is computed? The second is a question about mechanisms: How does a VP get omitted in that representation in some (but not all) languages? As far as we can tell, accounts sometimes address the first question obliquely, so where necessary, we will try to spell out what the underlying assumptions might be. Where appropriate, we also point out how these accounts handle cross-linguistic variation, although we note that not all of them were developed to deal with this issue.

After reviewing these questions, we discuss the perception of acceptability in Section 2.2: what aspect of that representation makes a missing VP sentence relatively acceptable, even though the sentence is ungrammatical?

2.1 Why is a VP omitted?

2.1.1 Structural forgetting

---

[2] We therefore set aside accounts that attribute the illusion to prosodic factors (e.g. Fodor et al. 2017). Briefly, these accounts argue that in English, at least, the acceptability of a sentence is partly dependent on prosody. A grammatical DCE sentence typically has a prosodic structure that is unbalanced or list-like, reducing its acceptability. It is argued that one way to improve the sentence's prosody is by omitting a constituent, like a VP. While this account provides insight into why grammatical DCE sentences are unacceptable, it does not directly address why the illusion occurs with a missing intermediate VP, nor why the illusion varies across languages.

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

For illustration, we consider how the parser might process a DCE sentence like (1). The first NP, *the novel*, triggers a prediction for a verb, which is 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. To reduce the memory load, the parser forgets the prediction with the highest memory cost. In Gibson & Thomas's theory of memory load, this is the prediction for a VP triggered by the second NP.[3]

Gibson & Thomas further demonstrate experimentally that in acceptable missing VP sentences, it is indeed the VP associated with the second NP that goes missing. Starting with a grammatical DCE sentence (3a), they systematically omit each of the three VPs to produce the missing VP conditions in (3b-d).

The logic of the manipulation goes as follows. Taking (3b) as an example, if it were the case that the parser forgets the VP prediction triggered by the third NP *the publishing company*, then the VP *had typed quickly* should be paired with *the horror author*, with the NP as its subject. Similarly, the VP *was banned by the local library* should be paired with *the novel*. These pairings are plausible, so (3b) should be acceptable, exhibiting the illusion.

In contrast, if the forgotten VP prediction is actually the one associated with the second NP *the horror author*, then in (3b) the VP *had typed quickly* should get paired with the third NP *the publishing company*, with this NP as the subject. Since this pairing is implausible, (3b) should be rated worse. In fact, under this hypothesis, one would expect (3c) to exhibit the illusion instead. For (3c), this hypothesis predicts that *the publishing company* should be interpreted as the subject of the VP *had recently fired*, while the NP *the novel* should be interpreted as the subject of the VP *was banned by the local library*. These pairings are plausible.

Gibson & Thomas's results show that sentences like (3c) show the missing VP illusion; sentences like (3b) and (3d) received worse ratings. To be clear, though, there was no positive evidence that the missing VP sentences in (3c) were more acceptable than their grammatical counterparts in (3a); both conditions turned out to receive very similar ratings. That said, the absence of a clear acceptability penalty for the ungrammatical missing VP sentences can be reasonably taken as evidence for 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].

---

[3] Gibson & Thomas also raise the possibility that it is the "word-string" whose predictions are costly that gets forgotten (pp. 241—242). Put differently, the parser should forget 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.

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] ----.

The structural forgetting hypothesis attributes the missing VP effect to general limitations on cognitive (specifically, memory) resources. To the extent that these limitations are invariant across individuals, it also predicts that all languages should exhibit the missing VP illusion. We further note that this hypothesis 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). More specifically, Christiansen & Chater argue that a connectionist network that treats sentences as a sequence of nouns and verbs is an adequate parsing model, by showing that it can model certain behavioral results related to the processing of embedded clauses, such as the missing VP illusion in English. That said, this model does not predict how nouns and verbs are related to each other. Because of this, one limitation of Christiansen & Chater's account is that it does not go as far as to predict that it is the middle VP that goes missing, unlike Gibson & Thomas's account.

2.1.2 Language experience

However, experimental evidence suggests that not all languages exhibit a missing VP illusion to the same extent, contrary to the prediction of the structural forgetting hypothesis. Reading time experiments indicate that string-equivalent German and Dutch sentences do not elicit the illusion (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). Frank & Ernst (2019) also provide offline acceptability and comprehensibility judgment ratings from Dutch showing the absence of the illusion.

To explain cross-linguistic variation, Vasishth, Frank, and colleagues have put forward a language experience hypothesis. 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 and sequences of verbs are 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]

---

[4] One question that arises is just what kind of experience German and Dutch speakers have with processing sequences of verbs: for instance, how many of these sequences consist of three verbs, as in grammatical DCE sentences? As far as we know, there is no quantitative study of this question in the language experience literature, although Frank et al. (2016:568) assert that three-verb sequences are more common in German and Dutch.

To be sure, as presented so far, this hypothesis is not a full theory of the illusion in the same way that the structural forgetting hypothesis is. As far as we can tell, researchers have fleshed out the hypothesis in two distinct ways.

The first is to integrate it with the structural forgetting hypothesis, so that the structural forgetting hypothesis can accommodate the German and Dutch facts. For instance, Vasishth et al. (2010) suggest that German and Dutch speakers' experience with processing verb-final structures might condition working memory, so that 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 the prediction for a verb (or VP), even in the event of high memory load. Similar intuitions are present in the computational modeling accounts 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). Here, it is assumed implicitly that sentences can be represented lineally as sequences of nouns and verbs (and other syntactic categories) (see also Christiansen & Chater 1999; Christiansen & MacDonald 2009). The verb-finality of German and Dutch means that sequences of three verbs are more common in these languages than in English (Frank et al. 2016). Consequently, German and Dutch speakers, sensitive to the statistics of their languages, find the presence of three consecutive verbs in grammatical DCE sentences unsurprising, which leads them to read these verbs more quickly and assign higher acceptability ratings to grammatical DCE sentences. English speakers, on the other hand, encounter such sequences of verbs less frequently, and so find missing VP sentences easier. However, as Frank & Ernst (2019) and Futrell et al. (2020) note, this alternative approach is a fairly weak theory of the illusion: it predicts that an English DCE sentence should become more acceptable as long as any VP is omitted, a prediction inconsistent with Gibson & Thomas's results.

2.1.3 Interference

Häussler & Bader (2015) and Bader (2016) provide alternative explanations of Gibson & Thomas's results by appealing 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 constructs a syntactic representation, attaching words into this representation as they appear. Both accounts also assume that the parser can successfully attach the first verb to the VP in the lower relative clause. Successful attachment in this case is argued to be a recency effect: 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. For a language like English, Häussler & Bader (2015) argue that the main clause position is more prominent: the main clause was the first clause created by the parser and therefore benefits from a primacy effect. Its prominence makes it more likely that the second verb is incorrectly attached there, leaving the relative clause without a verb.

Bader (2016) presents a different parsing scenario that can explain why the strength of the illusion differs between English and string-equivalent German (and Dutch) sentences. German has verb-second word order in main clauses, so the two attachment sites differ syntactically: for Bader, the relative clause site is inside a verb-final VP, while the main clause site is actually a verb-second

position inside CP, a structurally high functional projection. This structural difference helps the parser 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 and wrong attachment to the main clause site is more likely. Although Bader does not go into details, this proposal may be compatible with a cue-based retrieval system, in which structural features 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, say, 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 a verb is missing there. He also suggests that processing load might play a role. He points out that in these sentences, the subject is structurally high, and argues that this position is associated with a lower processing load, freeing up resources that improve parsing accuracy. Finally, for thoroughness, we note, following Häussler & Bader (2015), that primacy effects might provide yet another explanation: the attachment site for a verb in the main clause was created early, and is therefore 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 explicit discussion about why missing VP sentences in languages like English are perceived as more acceptable, despite their ungrammaticality. In general, existing accounts seem to feature the idea that the parser, when parsing a missing VP sentence, comes to hold certain expectations about the VPs in the sentence. These expectations are then borne out, producing a sense of completeness.

We can distinguish between two variants of this idea, based on the representation they assume for missing VP sentences. In the first variant, implied in an account like Gibson & Thomas's, sentences are represented as a set of constituents that are related to each other (although it is less clear how these relations are represented). The parser predicts upcoming constituents based on the constituents it has observed so far. Because of memory limitations, when it processes a DCE sentence, the parser comes to hold predictions for only two VPs, not three. Since the parser observes only two VPs, these predictions are borne out. One might assume that the bearing out of these predictions produces a sense of completeness and acceptability. This idea is also readily found in various computational models of the illusion (Christiansen & Chater 1999; Christiansen & MacDonald 2009; Engelmann & Vasishth 2009; Futrell & Levy 2017; Futrell et al. 2020). Setting aside differences in model architecture and theoretical assumptions, these models are shown to parse DCE sentences as if they are sequences of three nouns followed by two verbs.

In contrast, the interference accounts of Häussler & Bader (2015) and Bader (2016) hold that the parser builds a single syntactic structure with nodes for upcoming VPs. It is implicitly assumed that the parser expects to fill all nodes, even though the parser actually ends up leaving the intermediate VP node empty. According to Häussler & Bader (2015), the parser is more likely to overlook the fact that this node is empty, because it receives no primacy or recency advantage for verb attachment relative to the other VP nodes. In Bader's (2016) account, the empty VP node is said to be harder to detect, because it is more easily confused with the other structurally similar (but filled) VP nodes and perhaps because of processing load. Regardless of why detection is

difficult, the net result is that the parser mistakenly concludes that all three VP nodes have been filled and the sentence is complete.[5]

Here, we suggest an alternative, non-structural explanation of the relative acceptability of missing VP sentences. This hypothesis is based on the premise that one of the core objectives in parsing a sentence is to establish thematic relations between the referents and predicates denoted by its NPs and verbs: who did what to whom, so to speak. Setting aside why and how the VP for the second NP is omitted, we suggest the acceptability is the result of the parser successfully building thematic relations between all the referents and predicates in such a sentence, even though a VP is missing. Because there is no referent or predicate that 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*. There are a total of three NPs, each representing a referent, and two VPs, each representing a predicate. Informal intuitions, supported by Gibson & Thomas's results, suggest that all three NPs (more precisely, their referents) get related to some 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*, since *fired* is transitive, and one for *was banned by the local library*. These requirements are satisfied by the three subjects. We represent these thematic relations in Figure 1, using arrows to indicate which NPs are related to which VPs. It is easy to see that every referent and predicate is connected to something else. (For comparison purposes, we depict thematic relations for a grammatical DCE sentence in Figure 2.)
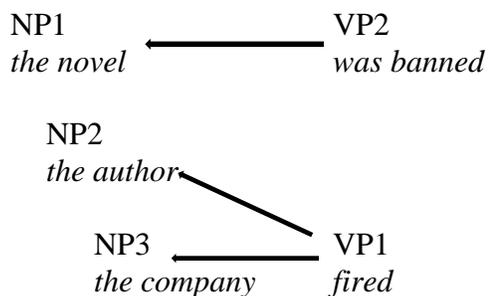

Figure 1: Thematic relations in missing VP sentence (3c) [NP1 The novel] that [NP2 the horror author] who [NP3 the publishing company] [VP1 had recently fired] [VP2 was banned by the local library].

---

[5] That said, because of the empty VP node, the resulting syntactic structure is ill-formed, even if the parser might happen to overlook this fact. This might explain why missing VP sentences usually receive relatively low acceptability ratings (see Gibson & Thomas 1999; Frank & Ernst 2019; and Experiment 2 below): perhaps the sense of completeness does not completely mask the percept of ill-formedness that is due to empty VP node.
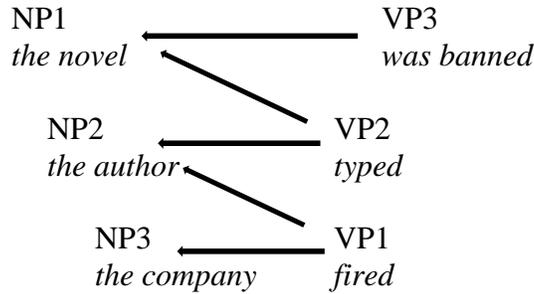
```
NP1  ←————————————  VP3
the novel                   was banned

NP2  ←————————————  VP2
the author ←                typed

NP3  ←————————————  VP1
the company                 fired
```

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 that arises here is what kind of representation is implicated in this thematic relation account. One possibility is that the relevant representation is syntactic, given existing proposals mapping syntactic relations to thematic relations and vice versa. For instance, under certain syntactic frameworks, theme arguments are NP complements of verbs, agent arguments are NP specifiers of verbs, etc. (e.g. Baker 1988; also Perlmutter & Postal 1984; Larson 1988).

However, adopting a purely syntactic approach also presents new problems. By definition, DCE sentences, including missing VP sentences, feature a doubly center-embedded relative clause, producing a single connected structure. In particular, NP1 and NP2 should each be syntactically related to two verbs: the verb of the relative clause modifying them and the verb of the clause that they are the subjects of, as reflected by the thematic relations in Figure 2. It is not immediately obvious what happens to these syntactic relations in an ungrammatical 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, non-syntactic representation in which the parser tracks the referents and predicates denoted by NPs and verbs. Thematic relations might be directly encoded between referents and predicates in such a representation, instead of being derived from a syntactic representation.

Of course, to the extent that shallow representations are involved, this suggestion appears to be inconsistent with a body of research in the psycholinguistics literature. There is substantial evidence suggesting that the parser incrementally builds a rich, connected syntactic representation, as revealed by the fact 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, for example, 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.). It seems plausible that under such circumstances, the parser finds it difficult to build and maintain a connected structure for interpretation. Consequently, it relies on a shallower

interpretive strategy, directly tracking how referents and predicates are linked to each other, so that it can produce a passable interpretation of the sentence.

Of course, even with this hypothesis about thematic relations in missing VP sentences, the fact remains that these sentences are syntactically ill-formed. All else being equal, we expect lower acceptability ratings for these sentences. This is not an issue that is unique to this account. In footnote 5, we pointed out the same issue for interference accounts. There, we observed that missing VP sentences typically do not receive high acceptability ratings, consistent with a scenario in which native speakers are sensitive to the syntactic ill-formedness of these sentences.

3 The present study

This article evaluates the above proposals about center-embedding illusions from the perspective of Mandarin Chinese. The word order properties of Mandarin conspire to produce DCE sentences characterized by a sequence of three transitive verbs followed by a sequence of NPs. We note that Mandarin appears to exhibit a missing NP illusion, not unlike how English and French show a missing VP illusion. We further highlight challenges in adapting existing hypotheses about missing VP illusions to account for the missing NP illusion.

We substantiate our arguments with a series of experiments. Experiment 1 demonstrates the missing NP illusion and determines how the transitive verbs and the NPs that follow them are thematically related. We show that in Mandarin missing NP sentences, the final NP is interpreted as an argument of all three verbs. In this respect, Mandarin 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 for English in Experiment 2, reproducing their findings.

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 3), 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
Kexuejia   hen   [$_{V1}$ xinshang][$_{V2}$ pipingguo][$_{V3}$ yanfa]   [$_{NP1}$ dianzichanpin] yizhenzi
scientist   very   admire   criticized   develop   electronics   a.while
de [$_{NP2}$ qiye]   yiliangci   de [$_{NP3}$ gongchengshi].
DE   corporation   a.few.times DE   engineer
'The scientist admires very much the engineer who criticized a few times the corporation that has developed electronics for a while.'

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
    Kexuejia   hen   [$_{V1}$ xinshang][$_{V2}$ pipingguo][$_{V3}$ yanfa]   [$_{NP1}$ dianzichanpin] yizhenzi
    scientist very   admire   criticized   develop   electronics   a.while
    de [$_{NP2}$ qiye].
    DE   corporation
    A possible "translation": 'The scientist admires very much ---- criticized the corporation that
    has developed electronics for a while.'

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, since Mandarin allows null objects (6a) and "headless" relative clauses where the head NP is omitted (6b). There are two reasons why it is inappropriate to assimilate missing NP sentences with either construction. First, in both constructions, the omitted NP is understood to refer to some entity that is salient in the discourse. These sentences elicit a specific percept of ill-formedness when presented out of context. In contrast, missing NP sentences do not; they are relatively acceptable even out of context.

(6) Null object and headless relative clauses infelicitous without appropriate context
    a. #Kexuejia   pipingguo __ (yiliangci).
       scientist   criticized   a.few.times
       'The scientist criticized [it/them/her/…] a few times.'
    b. #Kexuejia   hen   xinshang [$_{RC}$ pipingguo qiye   de] __.
       scientist   very admire   criticized corporation DE
       'The scientist admires very much the one [i.e. engineer/person/…] that criticized the
       corporation.'

Second, missing NP sentences can be distinguished syntactically from both constructions. In null object sentences, verbs usually appear at the end of a clause, unless there is adverbial material present (6a). In contrast, in missing NP sentences, none of the three verbs appear in such a position. Both V1 and V2 immediately 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 a NP and a *de*. In contrast, only a 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 are characterized by a sequence of three verbs and two NPs, while missing VP sentences are characterized by a sequence of three NPs and two VPs. One might therefore expect it to be straightforward to adapt the hypotheses proposed for the missing VP illusion to account for the missing NP illusion. However, more careful consideration suggests otherwise.

We begin with the challenges for the structural forgetting account (Gibson & Thomas 1999) and interference account (Häussler & Bader 2015; Bader 2016). Briefly, the mechanisms proposed yield very different outcomes for thematic relations in English missing VP sentences and Mandarin missing NP sentences, raising questions as to why the Mandarin sentences should be relatively acceptable.

We first present the intuition behind this argument. In English missing VP sentences, the parser is claimed to pair the two VPs with the first and third NPs, having forgotten or overlooked the structural predictions (for a VP) associated with the second NP. Despite the absence of a VP, the parser is still able to establish thematic relations between the three NPs and the two VPs. The three NPs are each connected to one VP or the other, while the two VPs require a total of three arguments, a requirement collectively satisfied by the three NPs.

By analogy, in Mandarin missing NP sentences, the parser should also pair the two NPs (the analog of English VPs) with the first and third verbs, having also forgotten or overlooked the predictions for a NP associated with the second verb. Here, however, we expect the absence of a NP to cause a more severe disruption in terms of thematic relations, which should have consequences for 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 for the verbs; we argue below that there is a distinct possibility that V2 would be "orphaned" thematically, without agent and theme arguments.

More specifically, suppose that a missing NP sentence like (5) is parsed in the following manner. First, the parser observes a NP, *scientist*, and then V1 *admire*, and interprets *scientist* as V1's subject and experiencer argument. Observing V1 further triggers a prediction for an NP object, its theme argument. Next, when V2 *criticized* is seen rather than V1's object, V2 is recognized as the main verb of a subject relative clause that modifies V1's object. V2 then triggers predictions for a NP object and a *de*, such that the object is to be interpreted as V2's theme argument, while the NP after *de* is to be interpreted as V2's agent argument. Lastly, when V3 *develop* is seen rather than V2's object, it is recognized as the main verb of another subject relative clause. The parser again predicts a NP object and *de*, such that this object is to be interpreted as V3's theme and the NP after *de* as V3's agent.

Both structural forgetting and interference hypotheses claim that the parser should forget or overlook the predictions associated with V2 *criticized* (but not V2 itself). Consequently, the parser should end up pairing the first post-verb NP *electronics* and *de* with V3 *develop*, and the second NP *corporation* with V1 *admire*. As the NP that appears right after the relative clause containing V3, *corporation* should also get interpreted as V3's agent.

Figure 3 sums up the thematic relations built in this scenario. The three NPs – the main clause subject *scientist*, NP1 *electronics* and NP2 *corporation* – are all linked to verbs. As for the three verbs, only two have arguments: V1 *admire* and V3 *develop*. V2 *criticized* is "orphaned," without an agent or a theme. If missing VP sentences are acceptable because the parser is able to build thematic relations between predicates and referents, 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.)
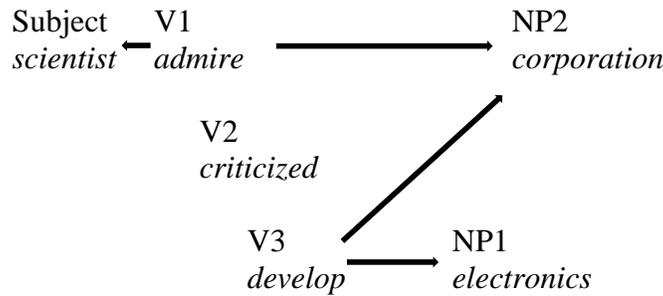
Subject   V1   NP2
scientist *admire*   *corporation*

        V2
        criticized

        V3   NP1
        develop   *electronics*

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

Subject   V1   NP3
scientist *admire*   *engineer*

        V2   NP2
        criticized *corporation*

        V3   NP1
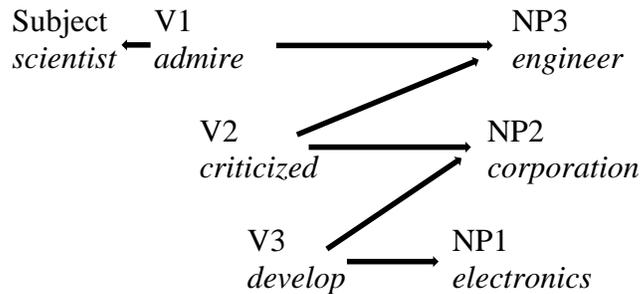   *develop*   *electronics*

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

For thoroughness, we note that one could accommodate the missing NP illusion by adopting a "shallower" approach in which sentences are represented as sequences of syntactic categories, without deeper commitments about thematic relations. This approach is often implicitly present in computational models of the missing VP illusion (e.g. Christiansen & Chater 1999; Frank et al. 2016; Futrell et al. 2020). As mentioned in Section 2.1.2, in these analyses, the parser predicts lexical items based on what it has seen so far. Upon processing V3 in a Mandarin DCE sentence, the parser might expect to see only a noun, a *de*, followed by another noun. These predictions are borne out in missing NP sentences, which could explain their relative acceptability.

While these analyses are logically coherent, they make for much weaker theories of the illusion. These analyses imply that the acceptability of DCE and missing VP sentences is simply a function of sequences of constituents. Compared to structural forgetting and interference accounts, there is relatively little they have to say about the finding that English speakers systematically interpret the two VPs in a missing VP sentence as taking the first and third NPs as their subjects. Consequently, the "shallower" approach does not give insight into how these parses or interpretations contribute to percepts of acceptability.

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. Here, however, we point out that 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 would yield the prediction that Mandarin should lack such an illusion. (Similar remarks hold for alternative implementations of this

hypothesis, where sentences are encoded as sequences of syntactic categories, as discussed 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 and all complements and modifiers are marked overtly with *de*. Consequently, sequences of nouns, whether interleaved with *de* or not, are common: 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 clause (7c) or a relative clause (7d). In fact, subject relative clauses, which correlate with noun sequences (as in (7d)), are reported to occur more frequently in naturalistic contexts than object relative clauses in Mandarin (Hsiao & Gibson 2003; Vasishth et al. 2013).

(7) Noun sequences in Mandarin (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. [gongchengshi piping   [$_{NP}$ kexuejia]] de   [$_{NP}$ yaoyan]
          engineer   criticize   scientist DE   rumor
        'the rumor that the engineer criticized the scientist'
    d. [piping   [$_{NP}$ kexuejia]] de [$_{NP}$ gongchengshi]
          criticize   scientist DE   engineer
        'the engineer who criticized the scientist'

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. As a result, Mandarin speakers should also be relatively familiar with processing long dependencies between a verb and the head of its object. The net effect is that grammatical DCE sentences, which feature long verb-object dependencies and sequences of nouns, should be easier to process and the missing NP illusion less likely.

Of course, the points made above assume that the missing NP illusion is robust 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 that was designed to validate these assumptions. As a preview, the results suggest that missing NP sentences are indeed relatively acceptable, but no verb is orphaned. Instead, we find evidence that speakers can and do interpret V2 as taking NP2 as an argument.

4 Experiment 1

4.1 Participants

Participants were 42 native speakers of Mandarin Chinese from mainland China, Taiwan, or Singapore, recruited from the University of Maryland community. They were compensated US$6 for their time.

4.2 Materials

A set of 24 DCE sentences were constructed, each with six 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 related 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 were variants of each other.

In addition, 72 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. 48 of the fillers were acceptable sentences and 24 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 DCE sentences, grammatical or otherwise, relatively unacceptable compared to the acceptable fillers.

Because the regions that participants came from use either simplified or traditional Chinese characters, experimental materials were prepared in both orthographic variants. Participants could choose which version they preferred.

In all the relative clauses, a duration or frequency adverb appeared between the object and the clause-final particle *de*. In (4) (repeated below as (8)), the duration adverb *yizhenzi* "a while" appears between NP1 *dianzichanpin* "electronics" and *de*, and the frequency adverb *yiliangci* "a few times" between NP2 *qiye* "corporation" and *de*.

(8) Mandarin Chinese DCE sentence
Kexuejia   hen   [$_{V1}$ xinshang][$_{V2}$ pipingguo] [$_{V3}$ yanfa]   [$_{NP1}$ dianzichanpin]   yizhenzi
scientist   very   admire   criticized   develop   electronics   a.while
de [$_{NP2}$ qiye]   yiliangci   de [$_{NP3}$ gongchengshi].
DE   corporation   a.few.times   DE   engineer
'The scientist admires very much the engineer who criticized a few times the corporation that has developed electronics for a while.'

The adverbs served to disambiguate the reading of *de*. Without an adverb, *de* in this position – between two nouns like *qiye* "corporation" and *gongchengshi* "engineer" – is in principle ambiguous. As mentioned, *de* can mark the end of a relative clause, in which case *qiye* is inside the relative clause and *gongchengshi* is the head of the relative clause. Alternatively, *de* can be understood as a possessive marker that marks the previous NP as a possessor and the following NP as the possessum (9), so *qiye* and *gongchengshi* form a single complex NP. The adverb's presence blocks the possession reading.

(9) qiye   de   gongchengshi
corporation   DE   engineer
'(the) corporation's engineer'

15

DCE sentences like (8) were contrasted with plausible missing NP variants like (10). Care was taken in the selection of lexical items to maximize the plausibility of these sentences. For instance, (10) is plausible in the sense that the second object NP, *qiye* "corporation" is a plausible theme argument for either V1, *hen xinshang* "admire very much," or V2, *pipingguo* "criticized", and a plausible agent (appearing as a relativized subject) of V3 *yanfa* "develop" or V2 *pipingguo*. By ensuring that NP2 was plausible in all of these roles, we avoided prejudging which verbs speakers choose to interpret it with in a missing NP illusion.

(10) Mandarin Chinese plausible missing NP sentence

Kexuejia   hen   [$_{V1}$ xinshang] [$_{V2}$ pipingguo] [$_{V3}$ yanfa]   [$_{NP1}$ dianzichanpin] yizhenzi
scientist   very   admire   criticized   develop   electronics   a.while
de   [$_{NP2}$ qiye].
DE   corporation
'The scientist admires very much ---- criticized the corporation that has developed electronics for a while.'

The verbs in each sentence 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. Without controlling for verb semantics and morphology, speakers might find it easier to coerce a reading of a missing NP sentence where two of the three verbs, such as V1 "admire" and V2 "envy" in (11), are conjoined with each other. A conjunction analysis would then produce a grammatical sentence with only one level of embedding, as illustrated by the English translation in (11a).

(11) Possible parses of missing NP sentence

Kexuejia   hen   [$_{V1}$ xinshang] [$_{V2}$ xianmu] [$_{V3}$ yanfa] [$_{NP1}$ dianzichanpin] yizhenzi
scientist   very   admire   envy   develop   electronics   a while
de   [$_{NP2}$ qiye].
DE   corporation
(a) Grammatical single center-embedding parse, to be avoided: 'The scientist admires very much *and* envies the corporation that has developed electronics for a while.'
(b) Ungrammatical DCE parse (if there is one): 'The scientist admires very much ---- envies the corporation that has developed electronics for a while.'

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 verb the NP was interpreted with.

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 (12a). While we do not give translations for sentences like (12a) because doing so would presuppose a parse, the intuition for this condition is as follows: if speakers interpret NP2 *qiye* 'corporation' as the theme argument of V1 *aishangle* 'fell in love with,' they should assign a lower acceptability rating to the sentence, since one does not usually fall in love with corporations. Likewise, in the second condition, V2 was chosen so that NP2 would

be an inappropriate theme argument (12b), 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 (12c): corporations do not serve as university presidents. For ease of reference, we call these the "implausible V1 object," "implausible V2 object," and "implausible V3 subject" conditions.

(12) Manipulations involving verbs and NP2
    a. Implausible V1 object (*qiye* "corporation" is an implausible object for V1 *aishangle* "fell in love with")
        Kexuejia   [$_{V1}$ **aishangle**]   [$_{V2}$ pipingguo] [$_{V3}$ yanfa]   [$_{NP1}$ dianzichanpin]
        scientist   fell.in.love.with   criticize   develop   electronics
        yizhenzi de [$_{NP2}$ **qiye**].
        a.while DE   corporation
    b. Implausible V2 object (*qiye* "corporation" is an implausible object for V2 *aishangle* "fell in love")
        Kexuejia hen [$_{V1}$ xinshang][$_{V2}$ **aishangle**]   [$_{V3}$ yanfa]   [$_{NP1}$ dianzichanpin]
        scientist very   admire   fell.in.love.with develop   electronics
        yizhenzi de [$_{NP2}$ **qiye**].
        a.while DE   corporation
    c. Implausible V3 subject (*qiye* "corporation" is an implausible subject for *danren daxue-xiaozhang yizhenzi* "to serve as a university president for a while")
        Kexuejia hen [$_{V1}$ xinshang][$_{V2}$ pipingguo][$_{V3}$ **danren**] [$_{NP1}$ **daxue-xiaozhang**]
        scientist very   admire   criticized   serve.as   university-president
        yizhenzi de [$_{NP2}$ **qiye**].
        a.while DE   corporation

        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 (13). We refer to this as the "implausible V3 object" condition.

(13) Implausible V3 object (*jiating* "family" is an implausible object for V3 *yanfa* "develop")
    Kexuejia hen [$_{V1}$ xinshang][$_{V2}$ pipingguo][$_{V3}$ **yanfa**] [$_{NP1}$ **jiating**] yizhenzi
    scientist very   admire   criticized   develop   family a.while
    de [$_{NP2}$ qiye].
    DE   corporation

4.3 Procedure

        Participants did the experiment in a lab setting. They were first asked whether they preferred to read in simplified or traditional Chinese, and the appropriate set of materials was then selected. 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 (*tongshun*), totally unacceptable" and "7" was "very coherent, totally acceptable." For the first four practice sentences, participants saw brief comments about each sentence and what kind of ratings to assign: for example, they were instructed that a sentence that violated Mandarin word order rules or described an unlikely scenario should be given a low rating. Each sentence was presented on the same screen as the rating scale. Participants rated the sentence using the

keyboard or by clicking on the rating scale on the screen. 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. The experiment usually took about 20—25 minutes to complete.

4.4 Data analysis

The sentences tested in this experiment were fairly long. To ensure that participants had taken the time to read a sentence before rating it, we excluded all responses that were provided in less than 3 seconds after the presentation of the sentence; this eliminated about 3.7% of responses, including test and filler sentences.

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), following Frank & Ernst's (2019) analysis. The model had conditions as a fixed effect and random intercepts for participant and items, to control for by-participant and by-item variance. The plausible missing NP condition was taken as the baseline for this analysis.

4.5 Results

Averaged ratings for the conditions are presented in Table 1. All conditions received acceptability ratings in the 3 to 4 range; no condition was particularly acceptable in absolute terms. Importantly, grammatical DCE sentences were rated as less acceptable than the plausible missing NP condition. In the mixed effects model, the estimate for the grammatical DCE condition, relative to the plausible missing NP baseline, was also significantly lower (estimate = -0.47, z = -2.37, p = 0.02).

Within the implausible missing NP conditions, the implausible V3 subject condition received relatively high ratings (an average of 3.66) compared to the other conditions. A post-hoc inspection of ratings suggests that this was driven by a subset of items, in which the NP is a non-stereotypical agent or subject (e.g. VP: *trains dancers*; NP: *painter*). This manipulation did not always work as well as intended, resulting in higher ratings.

Nevertheless, all the implausible missing NP conditions were less acceptable than the plausible missing NP condition, a point underscored by the statistical analysis (estimate for implausible V1 object condition: -0.78, z = -3.90, p < 0.01; estimate for implausible V2 object condition: -0.65, z = -3.30, p < 0.01; estimate for implausible V3 subject condition: -0.40, z = -2.01, p = 0.04; estimate for implausible V3 object condition: -1.12, z = -5.62, p < 0.01). In other words, acceptability ratings here were sensitive to plausibility manipulations.

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

Condition	Mean acceptability rating (standard error)
Grammatical DCE (8)	3.63 (0.15)
Plausible missing NP (10)	4.01 (0.15)
Implausible V1 object (12a)	3.32 (0.15)
Implausible V2 object (12b)	3.42 (0.15)
Implausible V3 subject (12c)	3.66 (0.16)
Implausible V3 object (13)	3.01 (0.14)

4.6 Discussion

Experiment 1 confirmed informal impressions that missing NP sentences elicit an illusion of grammaticality. The illusion is analogous to missing VP illusions reported for other languages, despite differences in word order and in the set of thematic relations constructed.

Another finding, based on the lower ratings for the implausible conditions, is that speakers can and do interpret the second post-verb NP as if it were an argument of all the three verbs. In particular, this NP appears to be understood as the theme argument (object) of V1 and V2. 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 argument, 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.

These results are arguably consistent with the thematic relations hypothesis, which attributes the acceptability observed for a missing VP/NP sentence to the parser establishing thematic relations between all referents 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 reasonably 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, the results here indicate clearly 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 6 (Experiment 3).

In sum, Experiment 1 provided empirical evidence that Mandarin exhibits a missing NP illusion that is clearly reminiscent of the missing VP illusion reported for English. In addition, the two post-verb NPs in missing NP sentences are interpreted as objects and subjects of the preceding verbs in a way that lacks a parallel in English missing VP sentences. However, these remarks presume the validity of existing descriptions of missing VP sentences. So that we have a stronger basis for comparison of Mandarin and English, we replicate Gibson & Thomas's experiment on the missing VP illusion with Experiment 2.

5 Experiment 2

Although Gibson & Thomas's experiment was presented as a test of the structural forgetting hypothesis, their manipulations are designed to determine whether English exhibits a missing VP illusion, and if so, which NPs serve as subjects of the two VPs. These are the questions that we are interested in addressing in this experiment.

5.1 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 (14). We follow Gibson & Thomas in labeling these conditions as grammatical DCE, missing VP1, missing VP2, and missing VP3 sentences.

(14) 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 determine whether speakers treat the third NP as the first VP's subject and whether they treat the first NP as the second VP's subject, assigning these NPs agent or experiencer interpretations. In the missing VP1 condition (14b), this would produce a plausibility violation: 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 (14d), 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. The 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 encounter as many acceptable sentences as unacceptable sentences, on the assumption that all four experimental conditions should be judged as 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.*

5.2 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.

5.3 Procedure

The procedure was largely the same as that of Experiment 1. We departed from Gibson & Thomas (1999) 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 the scale, so that "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.

5.4 Data analysis

The sentences tested were fairly long. To ensure that participants took the time to read the 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 5.3% of responses, including those for filler sentences.

As was the case for the first experiment, ratings for all conditions in this experiment were analyzed with a cumulative link mixed effects model, with conditions as a fixed effect and random intercepts for participant and items. Ratings for the missing VP2 condition were used as a baseline.

5.5 Results

Table 2 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). Statistically, this condition was significantly more acceptable than the missing VP1 condition (estimate = -1.29, z = -4.49, p < 0.01) and the missing VP3 condition (estimate = -1.40, z = -4.93, p < 0.01). In numerical terms, ratings for the missing VP2 condition were also higher than those for the grammatical DCE condition, but the difference is not significant (estimate = -0.26, z = -0.97, p = 0.33). In sum, we replicated Gibson & Thomas's key findings.

Table 2: Mean acceptability ratings for Experiment 2.
(1 = completely unacceptable, 7 = completely acceptable)

Condition	Mean acceptability rating (standard error)
Grammatical DCE (14a)	3.10 (0.17)
Missing VP1 (14b)	2.47 (0.16)
Missing VP2 (14c)	3.29 (0.17)
Missing VP3 (14d)	2.41 (0.15)

5.6 Discussion

Our replication confirms Gibson & Thomas's 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. The latter 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 treated as if it were the object of both the first and second verbs (the equivalent of the first and second NPs in English).

There was also no significant difference between the ungrammatical missing VP2 condition and the grammatical DCE condition. The null result is inconsistent with informal intuitions, where ungrammatical missing VP2 sentences are perceived as more acceptable than their grammatical counterparts. However, in this context, the null result is not unexpected. Experiment 2 used the same test sentences as Gibson & Thomas's study, which also did not detect an acceptability contrast. In addition, as mentioned by Gibson & Thomas, the off-line nature of the task allowed participants to re-read sentences; participants who did so might have noticed the ungrammaticality of missing VP2 sentences.

That being said, as we pointed out in Section 1, the null result still justifies calling the phenomenon a grammaticality illusion: the ungrammatical missing VP sentence here is perceived to be as acceptable as its grammatical counterpart, unlike most other ungrammatical sentences, which are easily judged to be worse.

To sum up, Experiment 2 confirms existing descriptions about the missing VP illusion in English. Together, Experiments 1 and 2 show that Mandarin and English speakers exceptionally perceive ungrammatical DCE sentences with missing NPs as VPs as no worse than their grammatical counterparts, 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 with DCE sentences, 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).

6 Experiment 3

The language experience hypothesis was proposed to explain why the missing VP illusion is weaker or non-existent in German and Dutch than in a language like English. According to this hypothesis, the verb-final syntax of German and Dutch means that speakers have more experience with maintaining predictions of verbs. The greater experience might result in more robust representations of these predictions, reducing the likelihood of structural forgetting. In addition, this fact about German and Dutch syntax means that sentences with consecutive verbs are generally more frequent, making the consecutive verbs found in DCE sentences less surprising and therefore more acceptable.

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.

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.

6.1 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 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. 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 separated from the verb by modifiers (if any). The focus on intransitive sentences appears to be 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.

For both languages, nouns are assumed to be followed by a modifier 50% of the time, of which 50% are relative clauses. For English, it is stipulated that 20% of relative clauses are verb-final object relative clauses, reflecting the fact that object relative clauses occur relatively infrequently 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, resulting in a more diverse set of sequences of varying grammaticality.

We extend the model to Mandarin 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. 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 actually 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.[6] However, we take that this model is not intended as an approximation of how actual differences in word order statistics can explain 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 specific differences in word order requirements, such as whether relative clauses are uniformly verb-final. One could see these parameter values as simplifying assumptions necessary for the demonstration of this point.

6.2 Model results

We first ran the model for English and German/Dutch. The goal was to replicate Futrell et al.'s results for these languages. 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. For presentation purposes, we report probabilities in terms of surprisal, in bits. 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, as shown in Table 3.

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 an 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 than the surprisal of observing the end-of-sentence symbol. The model bears out this prediction, as Table 3 shows.

---

[6] Bader (2016) shows that DCE-like sentences are vanishingly rare in German, counting only 163 sentences in a 11% sample of a 92 million-sentence corpus. Karlsson (2007), surveying corpora and naturally-occurring texts, reports no sentences with doubly-embedded object relative clauses in seven European languages, including English and German. We are not aware of any similar studies for Mandarin, but we know of no reason to suspect that DCE sentences are more frequent there.

Table 3: Noisy surprisal model results 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 due to subject relative clauses being 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 4 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 4: Mandarin surprisal difference and probability of subject relative clauses.

P(subject relative clauses | relative clause)	80% (see Table 3)	60%	40%	20%
Surprisal difference	-0.762	-0.616	-0.395	0.027

While the cross-linguistic distribution of the center-embedding illusion must be attributed to some kind of difference between German / Dutch, English, and Mandarin, this modeling exercise shows that it probably should not be understood in terms of the frequency of verb-final or noun-final structures, contrary to the language experience account.

7 General discussion

The present study looked at existing accounts of center-embedding illusions, evaluating them using novel experimental data from Mandarin Chinese. Mandarin presents an interesting test case: its word order means that illusory sentences involve sequences with a missing NP, instead of sequences with a missing VP, as is the case in existing studies of the illusion. The absence of a NP in principle affects how thematic relations are built within illusory Mandarin sentences, with consequences for hypotheses about why speakers perceive these ungrammatical sentences to be acceptable.

Experiment 1 confirmed that Mandarin speakers find missing NP sentences more acceptable than their grammatical counterparts, just as Experiment 2 showed that English speakers experience a missing VP illusion (replicating Gibson & Thomas 1999). However, both experiments also show that Mandarin and English speakers build rather different thematic relations between the nouns and verbs in these sentences.

Figure 5 summarizes the thematic relations in English missing VP sentences, based on Gibson & Thomas 1999/Experiment 2, and Mandarin missing NP sentences, based on Experiment 1. English missing VP sentences are "underconnected" relative to their Mandarin missing NP counterparts. More specifically, in English, the final VP is linked to only the first NP. In contrast, in Mandarin, it appears that the final NP is linked not only to the first verb, but also to the second and third verbs.
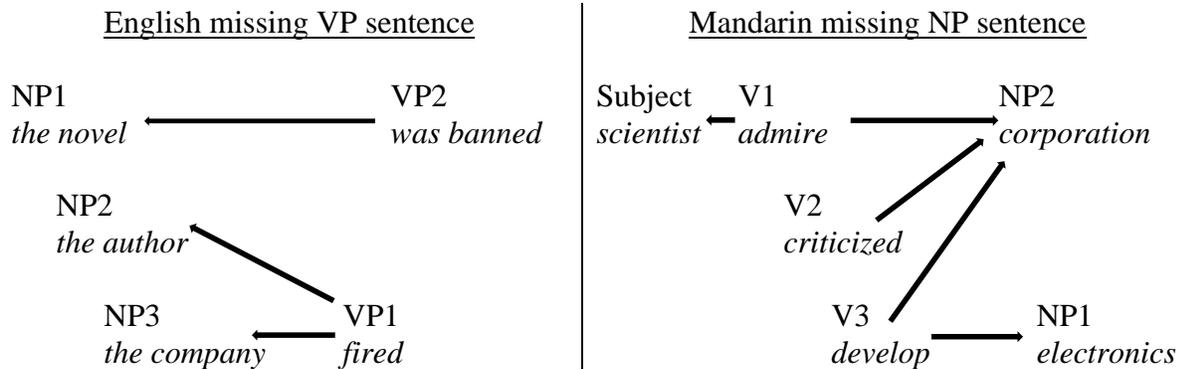
English missing VP sentence

NP1 *the novel* ⟵ VP2 *was banned*

NP2 *the author*

NP3 *the company* ⟵ VP1 *fired*

Mandarin missing NP sentence

Subject *scientist* ⟵ V1 *admire* ⟶ NP2 *corporation*

V2 *criticized*

V3 *develop* ⟶ NP1 *electronics*

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 referents 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 3, 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).
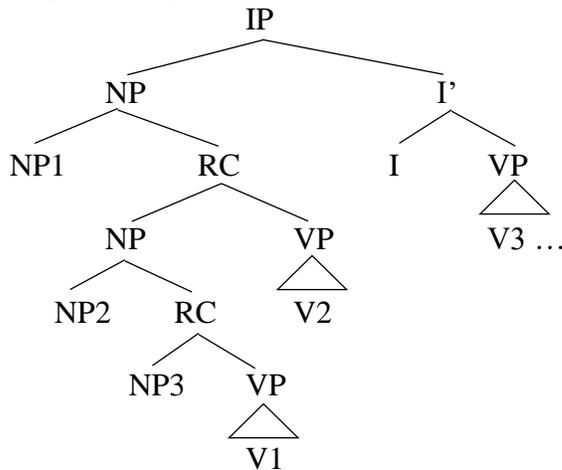
Here, we consider in greater detail an alternative account of cross-linguistic variation, namely, the interference account of Bader (2016), exploring how it can be adapted to accommodate the Mandarin data.

To recap, the key assumption in Bader's account is that German and English DCE sentences have different syntactic representations, even though they might be string-equivalent. In the German (and Dutch) examples where the illusion is weaker, the verb of the 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 (15a). The other two verbs, on the other hand, remain in VPs in their respective clauses. In contrast, English verbs all remain low in their respective VPs; they do not move to C (15b).

(15) a. German (adapted from Bader 2016, ex. 22)
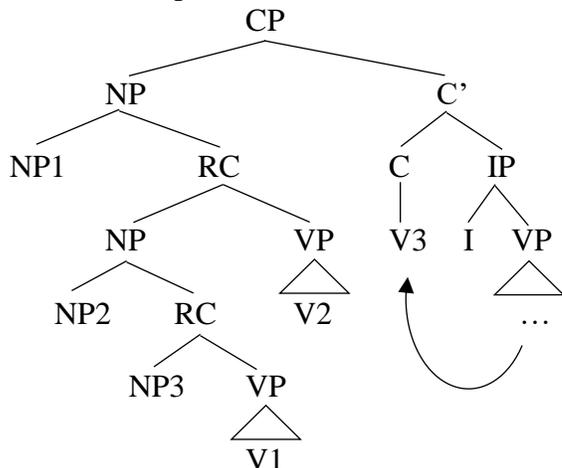
b. English (adapted from Bader 2016, ex. 17b)

```
                        IP
            ┌───────────┴───────────┐
           NP                       I'
      ┌─────┴─────┐            ┌─────┴─────┐
     NP1          RC           I          VP
             ┌─────┴─────┐                 △
            NP          VP                V3 …
        ┌────┴────┐     △
       NP2       RC     V2
             ┌────┴────┐
            NP3       VP
                      △
                      V1
```

According to Bader, the parser builds syntactic trees along the lines in (15), creating empty slots in anticipation of upcoming verbs. Bader argues that recency effects let the parser accurately integrate the first verb into the V1 slot in both languages. More specifically, this slot is created when the parser processes NP3 (or a relative pronoun immediately preceding NP3) and recognizes the presence of a relative clause. The parser then encounters the first verb. Correct attachment of the verb to the V1 slot is not a problem, since the clause and the slot were recently accessed and highly activated (see also Häussler & Bader 2015; Gibson & Thomas 1999).

The parser next observes the second verb, which in principle should be attached to the V2 slot and not the V3 slot. As (15) shows, in English, these two slots are both empty slots 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 German, however, these two slots are structurally different. V2 is inside a verb-final VP, while V3 is inside C. Consequently, the German parser is less likely to confuse the two slots and is more likely to attach the second verb correctly into V2.

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.

This account also needs to explain why, in German, the parser can proceed to successfully attach the third verb to the V3 slot. There are several suggestions present in the literature. For one, Bader (2016) suggests that the parser might find it easier to detect whether the V3 slot is empty, pointing to the structural distinctiveness of the V3 slot and potentially lower processing costs. Häussler & Bader (2015) argue that this slot enjoys a primacy advantage, since it is inside the main clause, the earliest predicted. The parser is therefore more likely to notice the fact that this clause is missing a verb.

Turning to Mandarin, we note that at an abstract level, the relevant syntactic facts resemble English more strongly than German, which explains why the illusion is present in Mandarin and English. Mandarin does not make any syntactic distinction between objects in main clauses and subordinate clauses (16). It therefore patterns like English, which does not make a distinction for verbs in main clauses and subordinate clauses, and unlike German, which makes a syntactic distinction.

For Mandarin, we assume that the parser can successfully integrate all words through the first *de* (including NP1). The parser then needs to attach the next word, the second noun, into the tree, ideally into the NP2 slot. As was the case for V2 and V3 in English, NP3 competes with NP2, since they are structurally similar: they are both NP slots dominated by another NP node. The Mandarin parser is thus likely to mis-attach the noun into the NP3 slot rather than the NP2 slot.

(16)   Mandarin
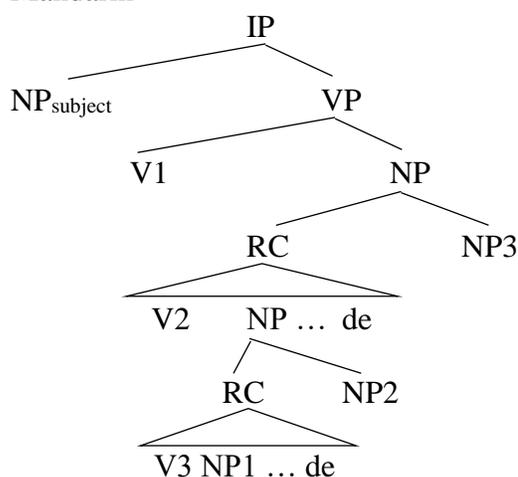
However, as presented so far, this analysis is empirically inadequate. As elaborated in previous sections, it predicts that only V1 and V3 have arguments; V2 might be "orphaned" thematically. This is inconsistent with the results of Experiment 1, which show that the second noun gets interpreted as a 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.

Here, 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. Even though interference is argued to cause the mis-attachment of the second verb, doing so 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). Repair therefore ensures that there are thematic relations between all referents and predicates, producing the same illusion of completeness.
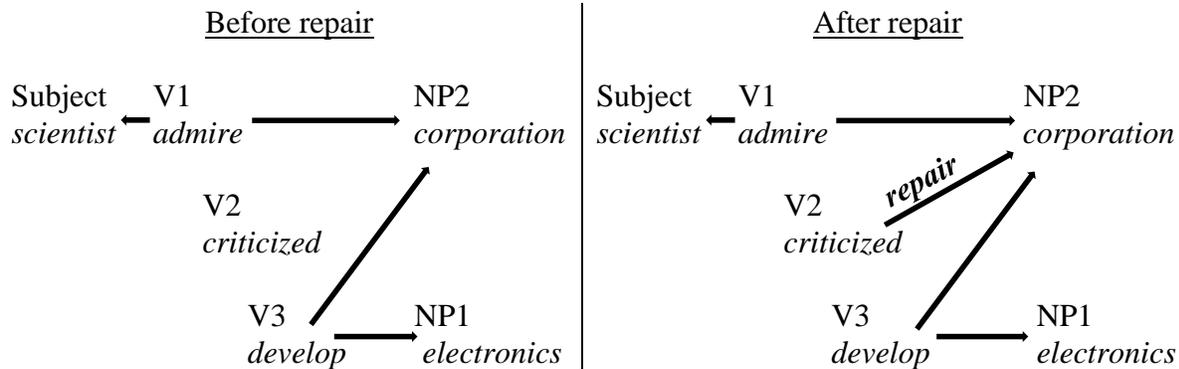
Before repair | After repair

Subject   V1   NP2   Subject   V1   NP2
scientist ← *admire* *corporation*   *scientist* ← *admire* *corporation*

V2   V2   *repair*
criticized   *criticized*

V3   NP1   V3   NP1
develop → *electronics*   *develop* → *electronics*

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 (16). 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, non-syntactic representation where thematic relations are encoded directly. 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.

8 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 assigned to these sentences do not follow easily from structural forgetting accounts like Gibson & Thomas's and interference accounts like Häussler and Bader's. The Mandarin data are also challenging for language experience hypotheses, which are intended to account for the cross-linguistic variation in center-embedding illusions (Vasishth et al. 2010; Frank et al. 2016; Futrell & Levy 2017; Frank & Ernst 2019; Futrell et al. 2020). These predict that Mandarin should lack the illusion.

Our experimental 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 referents 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.

Acknowledgments

References

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.

Baker, Mark C. 1988. *Incorporation: A theory of grammatical function changing*. Chicago, IL: University of Chicago Press.

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 & 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

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.

Larson, Richard K. 1988. On the double object construction. *Linguistic Inquiry* 19(3). 335–391.

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

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.

Perlmutter, David. M. & Paul M. Postal. 1984. Impersonal passives and some relational laws. In David M. Perlmutter and Carol G. Rosen (eds.), *Studies in Relational Grammar* 2, 126–170. Chicago, IL: The University of Chicago Press.

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

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