Journal of Memory and Language 84 (2015) 139–166

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Constructing covert dependencies—The case of Mandarin wh-in-situ dependency

Ming Xiang [a,*], SuiPing Wang [b], YanLing Cui [b]

[a] Language Processing Lab, Linguistics Department, University of Chicago, Chicago, IL 60615, United States
[b] School of Psychology, South China Normal University, Guangzhou, China

ARTICLE INFO

Article history:
Received 24 February 2014
revision received 3 April 2015

Keywords:
Wh-dependencies
Long distance dependencies
Covert dependency
Memory retrieval
Comprehension
Production

ABSTRACT

Wh-in-situ constructions in Mandarin Chinese, as opposed to their English counterparts that front wh-phrases to the beginning of the sentence, have the same word order as regular non-wh declaratives. We argue that despite their surface word order, processing wh-in-situ constructions involves constructing a covert non-local syntactic dependency between the in-situ wh-phrase and a higher scope position at a clause boundary, leading to behavioral patterns similar to those associated with the processing of overt dependencies. In two comprehension experiments, we showed that the process of linking an in-situ wh-phrase and its scope position induces a similarity-based memory interference effect if another clause boundary position intervenes. In addition, a set of sentence completion studies also showed that the production of wh-in-situ constructions is heavily modulated by the increased working memory burden that results from planning and maintaining a non-local dependency.

Introduction

Wh-dependencies have served as the flagship example for linguists and psycholinguists for examining the representations and processes involved in constructing linguistic dependencies. In a "wh-fronting" language, such as English, the wh-word is dislocated from its base position that bears the original theta role (e.g., agent, patient, etc.). In (1a) through (1c), the fronted wh-word is associated with the theme of the buying event (i.e., the object position after the verb *buy*). The surface position of the wh-word marks the semantic "scope" of the question. For instance, in (1a), the wh-word stands at the left edge of the sentence and signals a matrix question—a question that requires an answer; whereas in (1b) and (1c), the wh-word stays at the left edge of the embedded clause and only signals an embedded question—there is still a declarative sentence at the matrix level

(1) a. **What** did John buy __?
b. Mary knew **what** John bought __.
c. Mary wondered **what** John bought __.

The surface position of the wh-word in English (and other wh-fronting languages) provides useful cues for parsing. During left-to-right incremental parsing, encountering a wh-word is a clear signal to the parser to expect a non-local dependency. Indeed, one of the most robust findings in the parsing literature is that the parser prioritizes the forward expectation of a "gap" that is associated with the "wh-filler" (e.g., the active filler strategy, Frazier & Flores d'Arcais, 1989). Meanwhile, once the parser identifies a potential gap site, the features encoded earlier on the wh-filler guide memory retrieval of the filler item.

* Corresponding author. Fax: +1 773 834 0924.
E-mail address: mxiang@uchicago.edu (M. Xiang).

http://dx.doi.org/10.1016/j.jml.2015.05.006
0749-596X/

Previous research on non-local dependencies has focused heavily on examples like (1), in which the surface position of a "filler" explicitly cues the presence of a dependency, and the morpho-syntactic features on the filler also aid memory retrieval. We call such dependencies "overt" dependencies. It is well-known, however, that many languages, including Chinese, Korean, Hindi, etc., adopt a different strategy to form their wh-constructions. These are the so-called *wh-in-situ* languages: wh-questions in these languages have the same word order as their declarative counterparts, with the wh-words staying in their canonical theta positions. Examples of Chinese wh-in-situ questions are demonstrated in (2) below:

(2) a. Yuehan mai-le **shenme**?
John buy-perf **what**
What did John buy?

b. Mali zhidao Yuehan mai-le **shenme**.
Mary know John buy-perf **what**
Mary knew what John bought.

c. Mali xiangzhidao Yuehan mai-le **shenme**.
Mary wonder John buy-perf **what**
Mary wondered what John bought.

Despite the fact that the semantic scope of the question is not overtly specified for the sentences in (2)—i.e., the wh-words do not occupy their scope positions in the surface word order, the scope of these questions is clearly captured in the semantics of the sentences, as shown by the fact that these sentences receive exactly the same interpretations as their English counterparts in (1).

This paper examines the processing of wh-in-situ constructions. In the sections below, we will first lay out the basic grammatical assumptions about how the surface word order of a wh-in-situ question can be mapped onto a semantic interpretation shared by its wh-fronted English counterpart. As will become clear, the hypothesized grammatical representations underlying such constructions pose questions about whether and how the processing system postulates abstract non-local dependencies that do not have any morpho-phonological reflex on the surface word string—hence, their descriptor "covert". We approach the questions in two comprehension experiments by examining whether processing wh-in-situ dependencies triggers memory retrieval interference, which has been widely shown for the processing of overt non-local dependencies. Additionally, in a set of sentence completion studies, we also show that wh-in-situ constructions demonstrate production complexities analogous to those found for overt dependencies. Taken together, the current study lends support for the presence of covert dependencies in processing, and also for a unified mechanism that can account for the processing of typologically distinct non-local dependencies.

"Covert" dependencies in wh-in-situ constructions

A number of theoretical proposals have argued that, to map a Chinese wh-in-situ construction from its surface word order to its semantic interpretation, the in-situ wh-phrase needs to be associated with its scope position via an abstract syntactic dependency. Proposals differ in their specific technical implementations (Aoun & Li, 1993; Cheng, 1991, 2003; Huang, 1982; Tsai, 1994), but all of them postulate a non-local dependency between an in-situ wh-phrase and a silent structural position that is unmarked for any overt morpho-syntactic content. As shown schematically in (3), for a matrix question, the wh-phrase is associated with the left edge of the matrix clause (3a), and for an embedded question with the left edge of the embedded clause (3b). For descriptive convenience, we use "CP" to notate clauses, and "Q_{wh}" to notate a *silent* operator at the relevant clause edge position which performs the wh-interrogative function (i.e., resulting in the interpretation of a wh-question).

(3) a. $[_{CP1}\ Q_{whi} \ldots [_{CP2} \ldots [\ldots.wh_i \ldots.]]]$ Matrix Question

b. $[_{CP1} \ldots [_{CP2}\ Q_{whi} \ldots [\ldots.wh_i \ldots.]]]$ Embedded Question

The dependencies presented in (3) are "covert", because in contrast to their "overt" counterparts (e.g., English wh-questions), the left-hand side of the dependency is not marked by any morpho-syntactic cues in the surface word string. Two immediate questions arise. First, although the majority of the syntactic proposals on Chinese questions have assumed a covert dependency as above (modulo the differences in the specific syntactic implementations), there is very little experimental evidence to demonstrate that such a dependency is actually created during processing. And, second, if processing a wh-in-situ question indeed involves constructing a covert dependency, it is unclear whether the parser adopts the same computations and strategies as for overt dependencies. We see the representation and the processing questions above as two sides of the same coin—they mutually constrain and inform each other. In the current paper, by probing the comprehension mechanisms involved in understanding wh-in-situ questions, we also provide evidence for the underlying representations involved. Additionally, we will also discuss evidence from production that corroborates the comprehension results.

The comprehension of wh-in-situ constructions

Memory retrieval and memory interference in comprehension

In sentence comprehension, it is well established that non-local dependencies are subject to a distance effect: longer dependencies are generally more difficult to process (e.g., they have longer reading times, lower acceptability judgments, etc., Gibson, 1998; Lewis & Vasishth, 2005; Van Dyke & Lewis, 2003; Warren & Gibson, 2002). Both memory maintenance and memory interference factors have been proposed to account for such an effect. We focus on memory interference here (Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Van Dyke, 2002; Van Dyke & Johns, 2012; Gordon, Hendrick, & Johnson, 2001;

2004). Increasing the distance between a gap and its filler increases the likelihood that some intermediate material may introduce features shared with the retrieval target, and therefore makes the target more difficult to retrieve due to similarity interference.[1] The very fact that intervening material sharing features with the retrieval target can influence comprehension provides a strong diagnostic paradigm for detecting the presence of a non-local dependency. This is especially informative for examining wh-in-situ questions. If comprehending wh-in-situ questions indeed involves constructing an abstract non-local dependency, when comprehenders encounter the wh-in-situ phrase, a backwards retrieval process will be initiated to find the correct target position that the wh-phrase is associated with. And if, as the theoretical proposals hypothesize, the retrieval target is a higher clausal boundary position (i.e., a higher CP position), we would expect intermediate clausal boundaries to induce interference.

Memory retrieval cues

A core component of the memory retrieval process resides in a clear characterization of the possible retrieval cues. In previous comprehension studies, when the target of memory retrieval is a specific lexical item that bears certain overtly marked morpho-syntactic features (e.g., number, gender features, etc., on the target), intervening features of the same kind have been shown to induce interference (Bock & Eberhard, 1993; Bock & Miller, 1991; Pearlmutter et al., 1999; Wagers et al., 2009). Cue overload from conceptual features, i.e., features based on real-world knowledge, can also introduce interference effects (Van Dyke, 2007). The retrieval target for wh-in-situ questions, however, is a syntactic position at which the scope of a question is interpreted. The matrix clausal position (i.e., matrix CP) needs to be retrieved in the case of a matrix wh-in-situ question; and the embedded clausal position is retrieved for an embedded wh-in-situ question. In other words, the retrieval cues available to the parser will be configurational ones like "matrix CP" or "embedded CP". It is unclear whether configurational cues of this kind can be targeted by the parser in the same way as are other more explicit lexical features and induce similar interference effects. The impoverished retrieval cues for Chinese wh-in-situ questions also highlight some differences between different kinds of wh-in-situ languages. For example, a number of previous studies have examined the processing difficulty associated with in-situ wh-elements in Japanese (Aoshima, Phillips, & Weinberg, 2004; Miyamoto & Takahashi, 2002; Ueno & Kluender, 2009), demonstrating the presence of an active-filler strategy and length effect in Japanese questions. However, although a wh-phrase in Japanese does not need to leave its canonical theta position, it finds its scope position by associating itself with an overt scope marker attached to the verb at the appropriate position (*-no* or *-ka*). Overt morpho-syntactic cues in Japanese in-situ wh-questions, therefore, play a substantial role in guiding the retrieval of the correct target position. It remains to be seen how a retrieval target that is not overtly marked (as in Chinese) is processed.

Some initial evidence from Van Dyke and Lewis (2003) is informative in this regard. They showed that during the reanalysis of a mis-parsed garden-path sentence, the retrieval of the target attachment site was affected by intermediate constituents that shared similar structural properties with the target. Consider the following two sentences:

(4) a. High interference
The secretary **forgot** the student who **knew** the exam was important **was** standing in the hallway.

b. Low interference
The secretary **forgot** the student who was waiting for the exam **was** standing in the hallway.

In both conditions, the disambiguating verb "was" triggers the retrieval of a higher clause position at which the initial parsing error originated (i.e., the retrieval cue [+CP] guides the parser to find the position after "forgot"). But such retrieval is more difficult in the "high interference" condition (4a) than the "low interference" (4b), since the intervening verb "know" in (4a) also takes a clausal complement and therefore introduces an interfering [+CP] position.

Testing memory interference in the comprehension of in-situ wh-dependencies

Adopting the insight from Van Dyke and Lewis (2003), we hypothesize that intervening CP structures can introduce an interference effect when a higher CP position is being called for to construct a non-local dependency. Such an interference effect, if indeed confirmed, would provide strong evidence that an abstract covert dependency is instantiated in online processing. The crucial pair of comparison (among all the comparisons we will examine) is demonstrated below:

[1] In the current paper, we focus on the extra processing cost associated with interference. We note that interference sometimes creates facilitation effect (e.g. faster reading time), rather than additional cost, as in the cases of agreement attraction (Pearlmutter, Garnsey, & Bock, 1999; Wagers, Lau, & Phillips, 2009).

(5) Embedded wh-questions

a. Multi-CP

记者们	***想知道*** [$_{CP1}$	市长	***宣布*** [$_{CP2}$	市政府	严惩了	哪些官员]]
jizhemen	xiang-zhidao	shizhang	xuanbu	shizhengfu	yancheng-le	naxie-guanyuan
NP_1	V_1 [$_{CP1}$	NP_2	V_2 [$_{CP2}$	NP_3	V_3	NP_4]]
Reporter	**wonder**	mayor	**announce**	city-council	punish	**which officials**.

"The reporters wondered which officials the mayor announced that the city council punished."

b. Serial Verb

记者们	***想知道*** [$_{CP1}$	市长	***带领*** [$_{FP}$	市政府	严惩了	哪些官员]]
jizhemen	xiang-zhidao	shizhang	dailing	shizhengfu	yancheng-le	naxie-guanyuan
NP_1	V_1 [$_{CP1}$	NP_2	V_2 [$_{FP}$	NP_3	V_3	NP_4]]
Reporter	**wonder**	mayor	**lead**	city-council	punish	**which officials**.

"The reporters wondered which officials the mayor led the city council to punish."

The matrix verb (V1, *wonder*)[2] in (5a) obligatorily takes an interrogative complement.[3] In this particular case, it is an embedded wh-question ending with the in-situ wh-phrase *which officials* (NP4). The embedded clause (CP1) itself contains a predicate *announce* (V2) that takes another embedded clause (CP2). We call this condition a *Multi-CP* condition. At the critical region, i.e., the sentence final Wh-NP *which officials*, if interpreting the embedded in-situ question involves building a non-local dependency, the parser needs to retrieve the scope position that the wh-phrase is associated with, that is, the clause-initial position after the V1 *wonder*.[4] We notate the relevant retrieval cues as [+Q, +CP], with [+CP] representing a clause edge position and [+Q] representing a feature that, once associated with an appropriate in-situ wh-phrase, derives a wh-interrogative. The retrieval process guided by these retrieval cues, however, is likely to induce similarity-based interference in (5a), since the embedded verb V2 (*announce*) also introduces a CP complement. A Multi-CP sentence like (5a) is therefore a high interference condition, demanding extra processing effort.

What we call the "Serial Verb" condition, (5b), is maximally similar to (5a), but the crucial difference is that the embedded verb in (5b), "*lead*", does not take a full-fledged clausal complement (Grano, 2012). Note that the English counterpart of the same verb also does not take a full clause as its complement, but instead is used with an infinitival structure (e.g., *lead somebody to do something*, but not **lead that* ...). For descriptive convenience, we notate the structure following "*lead*" as an "*FP*" ("functional projection"), and distinguish it from the full clausal CP structure selected by the verb *announce* in (5a).[5] The sentence in (5b) serves as a critical comparison to (5a). These two conditions are matched on a number of dimensions, including overall length and the number of events (in terms of the number of verbs involved). But despite these similarities, we predict (5a) will be more costly than (5b), since only (5a) contains an intervening CP structure that will evoke additional interference cost.

The current study

We created eight conditions for the comprehension experiments. The first four were wh-in-situ conditions, shown in (6). The "Multi-CP" and "Serial verb" conditions in (6a) and (6b) are repeated from the examples in (5). In the "Adv" condition in (6c), the sentence is roughly matched with (6a) and (6b) for linear length,[6] through the addition of two adverbial phrases. And finally, the "Short" condition in (6d) is shorter in linear length

[2] Throughout the paper, we used "V" to represent predicates, including complex predicates in our stimuli. For example, V1 in (5a) is literally "want know", but we glossed it as "wonder".

[3] Just like in English, a verb like "wonder", although it obligatorily subcategorizes for an interrogative complement, can take either a wh-question complement (e.g. *John wondered what Bill bought.*) or a polar (yes–no) question complement (e.g. *John wondered whether Bill bought a book.*).

[4] It is important to note that the example in (5a) is unambiguous. First, it is impossible to have a matrix question reading like "Which official did the reporters wonder that the mayor announced that the city council punished?". Second, as an embedded question, the scope of the wh-in-situ phrase has to be associated with the matrix predicate V1 "wonder", not the predicate V2 "announce". That is, the sentence could not mean "The reporters wondered that the mayor announced which official the city council punished.". The obligatory association between the wh-in-situ phrase and the clause boundary after V1 is due to the fact that "xiangzhidao (wonder)" requires a [+Q] complement, which cannot be satisfied if the wh-phrase is associated with other scope positions. If we change the predicate "wonder" in (5a) to a predicate that can optionally allow, but does not obligatorily demand, a [+Q] complement, such as *hear*, *say*, *know*, etc., ("announce" belongs to this category as well), the sentence becomes three-way ambiguous: the wh-phrase could have the highest matrix scope (i.e. a direct question that requires an answer), or an embedded scope under either V1 or V2.

[5] Verbs like "lead" are also called "object control" verbs. We called them "serial verb" condition to be neutral in terminology. There is extensive discussion in Chinese syntax about the complement structure after object control verbs. Although the finite vs. non-finiteness distinction is a controversial one in Mandarin syntax (Hu, Pan, & Xu, 2001; Huang, 1989, 1994; Li, 1990), based on other diagnostics, such as whether an embedded aspectual marker receives a matrix or local interpretation, and whether inner-topicalization (topicalization to the edge of the embedded complement) is possible, a number of authors have convincingly advanced the proposal that whereas a regular verb like "say" or "announce" takes a CP (clause) complement, an object control verb only takes a complement that is structurally smaller than a full CP (Grano, 2012, 2014; Ernst & Wang, 1995; Paul, 2005a, 2005b).

[6] Compared to the "Short" conditions, the long conditions are all two words longer. But in terms of the number of characters, the Multi-CP conditions are on average 4.6 characters longer than the Short condition, the Serial Verb conditions are 4.5 characters longer (not different from the Multi-CP conditions, $p > .3$), and the Adv conditions are 5 characters longer, slightly more than the other two long conditions. ($p < .01$).

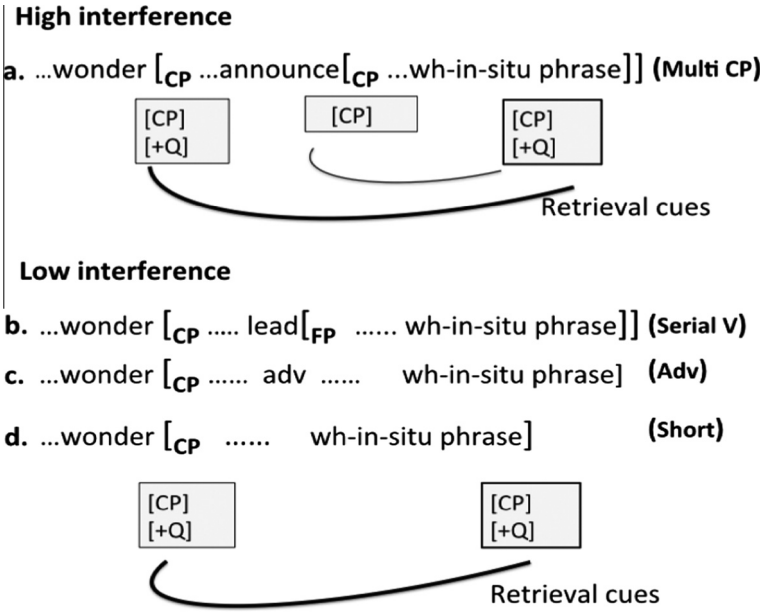


Fig. 1. Retrieval cues and interference effect for the high and low-interference WH-conditions.

compared to the other three. The critical Wh-NP (CW) is at the sentence final position due to the fact that Chinese is an SVO language and noun phrases are head-final, and therefore post-nominal material after the object noun phrase is uncommon.

Among these four WH-conditions, three are long, and one is short. We only predict an interference effect on the critical Wh-NP region for one of the long conditions—the Multi-CP condition—since this is the only condition that contains an intervening CP boundary. We schematically demonstrate

(6) WH-in-situ conditions

a. Multi-CP

记者们	***想知道*** [$_{CP1}$	市长	***宣布*** [$_{CP2}$	市政府	严惩了	哪些官员]]
jizhemen	xiang-zhidao	shizhang	xuanbu	shizhengfu yancheng-le		naxie-guanyuan
NP_1	V_1 [$_{CP1}$	NP_2	V_2 [$_{CP2}$	NP_3	V_3	NP_4]]
Reporter	**wonder**	mayor	**announce**	city-council punish		**which-CL official**.

"The reporters wondered which officials the mayor announced that the city council punished."

b. Serial Verb

记者们	***想知道*** [$_{CP1}$	市长	***带领*** [$_{FP}$	市政府	严惩了	哪些官员]]
jizhemen	xiang-zhidao	shizhang	dialing	shizhengfu yancheng-le		naxie-guanyuan
NP_1	V_1 [$_{CP1}$	NP_2	V_2 [$_{FP}$	NP_3	V_3	NP_4]]
Reporter	**wonder**	mayor	**lead**	city-council	punish	**which-CL official**.

"The reporters wondered which officials the mayor led the city council to punish."

c. Adv

记者们	***想知道*** [$_{CP1}$	市长	去年	夏天	严惩了	哪些官员]
jizhemen	wonder	shizhang	last-year	summer	yancheng-le	naxie-guanyuan
NP_1	V_1 [$_{CP1}$	NP_2	adv1	adv2	V_2	NP_3]
Reporter	**wonder**	mayor	**last summer**		punish	**which-CL official**.

"The reporters wondered which officials the mayor last summer punished."

d. Short

记者们	***想知道*** [$_{CP1}$	市长	严惩了	哪些官员]
jizhemen	wonder	shizhang	yancheng-le	naxie-guanyuan
NP_1	V_1 [$_{CP1}$	NP_2	V_2	NP_3]
Reporter	**wonder**	mayor	punish **which-CL official**.	

"The reporters wondered which officials the mayor punished."

the interference difference between these four conditions in Fig. 1. The other two long conditions, the Serial Verb and Adv conditions, serve to control for any processing complexity effect associated with factors other than interference. For example, compared to the Short condition, longer sentences may show increased processing difficulty due to an integration cost associated with complex events or mere linear length; but crucially, we expect to observe extra cost on the Multi-CP condition above and beyond the cost generated by any baseline complexity effects, as estimated by the comparison between the Multi-CP condition and the other long conditions.

In addition to the WH-conditions, four parallel declarative constructions were included as controls. An example is given in (7). These sentences have the same word order as the wh-sentences above, and the design of the conditions is also maximally similar, but with two modifications. First, we changed the matrix verb V1 from a predicate like *wonder* to a predicate like *be happy* (or *know*, *discover*, *hear*, etc.), which could take a declarative complement; and second, we changed the sentence final Wh-NP to a regular referential noun phrase.

with retrieval interference. The main purpose of the declarative conditions is to control for any baseline differences between different constructions, especially the difference between the Multi-CP construction and the other constructions. We predict that, if only the WH-multi-CP condition evokes additional interference cost, the processing cost difference between the WH-multi-CP condition and the other Wh-conditions should be larger than the baseline difference (if any) between the Declarative-Multi-CP condition and the other declarative conditions. That is, we expect an interaction between *Structure Type* (WH vs. Declaratives) and *Complexity* (Multi-CP, Serial Verb, Adv, and Short).

Our primary interest in the current paper is to assess the comprehension cost associated with retrieving a higher scope position for the in-situ wh-phrase. But memory retrieval is unlikely to be the only source of online comprehension complexity. Previous work has established that parsing complexity is at least partially driven by comprehenders' probabilistic belief about the possible syntactic outcome conditioned by the current sentence fragment (Hale, 2001, 2003; Levy, 2008). More generally speaking,

(7) Declaratives

a. Multi-CP

记者们	***很高兴*** [$_{CP1}$	市长	***宣布*** [$_{CP2}$	市政府	严惩了	那些官员]]
jizhemen	happy	shizhang	xuanbu	shizhengfu	yancheng-le	naxie-guanyuan
NP_1	V_1 [$_{CP1}$	NP_2	V_2 [$_{CP2}$	NP_3	V_3	NP_4]]
Reporter	**happy**	mayor	**announce**	city-council	punish	**those-CL official**.

"The reporters were happy that the mayor announced that the city council punished those officials."

b. Serial Verb

记者们	***很高兴*** [$_{CP1}$	市长	***带领*** [$_{FP}$	市政府	严惩了	那些官员]]
jizhemen	happy	shizhang	lead	shizhengfu	yancheng-le	naxie-guanyuan
NP_1	V_1 [$_{CP1}$	NP_2	V_2 [$_{FP}$	NP_3	V_3	NP_4]]
Reporter	**happy**	mayor	**lead**	city-council	punish	**those-CL official**.

"The reporters were happy that the mayor led the city council to punish those officials."

c. Adv

记者们	***很高兴*** [$_{CP1}$	市长	去年	夏天	严惩了	那些官员]
jizhemen	happy	shizhang	last-year	summer	yancheng-le	naxie-guanyuan
NP_1	V_1 [$_{CP1}$	NP_2	adv1	adv2	V_2	NP_3]
Reporter	**happy**	mayor	**last**	**summer**	punish	**those-CL official**.

"The reporters were happy that the mayor last summer punished those officials."

d. Short

记者们	***很高兴*** [$_{CP1}$	市长	严惩了	那些官员]
jizhemen	happy	shizhang	yancheng-le	naxie-guanyuan
NP_1	V_1 [$_{CP1}$	NP_2	V_2	NP_3]
Reporter	**happy**	mayor	punish	**those-CL official**.

"The reporters were happy that the mayor punished those officials."

Since these declarative sentences do not involve non-local dependencies, we predict no processing cost associated

sentence comprehension complexity has also been associated with production complexity, even though the exact

relationship between the two remains an issue under debate (MacDonald, 2013; Pickering & Garrod, 2013). Before our comprehension experiments, we therefore first conducted a series of sentence completion studies (Experiment 1) to quantify the structural frequency of the target structures that appeared in the comprehension experiments (Experiments 2 and 3). In these completion studies, sentence preambles with different sizes were given to participants to complete, providing us with information that help gauging incremental expectation for the target structures. The first three completion studies (Experiments 1a–c) mainly compared the completion difference between the Multi-CP and the Serial-Verb conditions (under both WH and Declarative constructions). In the last completion study (Experiment 1d), we compare the completion rate of the critical sentence-final Wh-NP for all four WH-conditions in (6). In addition to providing the structural norming information for the comprehension experiments, the results from these sentence completion studies also have more general implications about the production complexity of wh-in-situ dependencies. We will return to these general implications in the 'General discussion' section.

Experiment 1 – sentence completion studies

If a speaker is asked to complete the sentence fragment given in (8), what are his/her choices?

(8)

记者们	***想知道***	市长	***宣布*** _____
jizhemen	xiang-zhidao	shizhang	xuanbu
reporter	wonder	mayor	announce
N1	V1	N2	V2

Since the matrix predicate V1 is "wonder", which signals an interrogative complement, an obvious option is to continue the sentence into a wh-in-situ construction, as shown in (9a) and (9b). Both (9a) and (9b) are right branching structures, as illustrated by the structure diagrams.

(9) Right branching structure:

a.

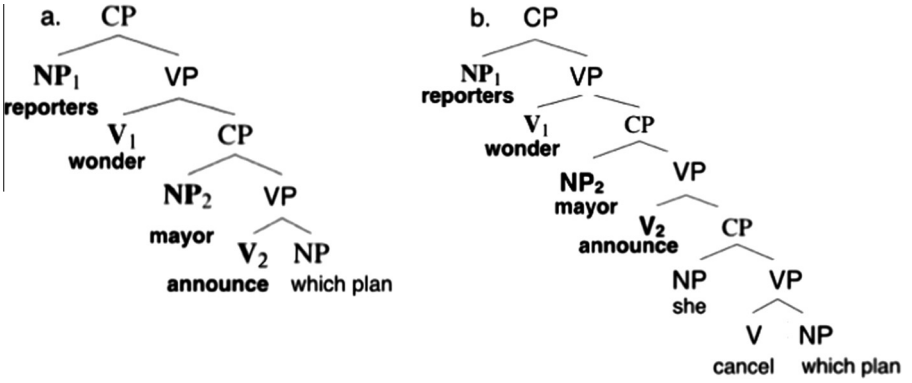
记者们	想知道	市长	宣布	了 哪个 计划
jizhemen	xiang-zhidao	shizhang	xuanbu	le nage jihua
reporters	wonder	mayor	announce	perf. which plan

"The reporters wondered which plan the mayor announced."

b.

记者们	想知道	市长	宣布	她 取消了 哪个 计划
jizhemen	xiang-zhidao	shizhang	xuanbu	ta quxiaole nage jihua
reporters	wonder	mayor	announce	she cancel-perf. which plan

"The reporters wondered which plan the mayor announced that she canceled."



An alternative, which both satisfies the requirement of "wonder" for an interrogative complement, yet at the same time does not evoke a wh-dependency, is to produce a "concealed question". We show some possible outcomes

below in (10). In these examples, syntactically speaking, the object of "wonder" is only a noun phrase (with a relative clause) instead of an actual interrogative clause; but semantically speaking, it conveys a question-like meaning. For example, a sentence beginning "*the reporters wondered about the reason that* . . ." could be paraphrased as "*the reporters wondered what the reason is that* . . .". There is good evidence, both theoretical and experimental, that the noun phrase object in a concealed question is semantically more complex than a regular noun phrase (Harris, Pylkkänen, McElree, & Frisson, 2008; Nathan, 2006; Romero, 2006).

(10) Relative clause structure:

a.

记者们	想知道	市长	宣布	的 那个 计划
jizhemen	xiangzhidao	shizhang	xuanbu	DE nage jihua
reporters	wonder	mayor	announce	DE that plan

"The reporters wondered about the plan that the mayor announced."

b.

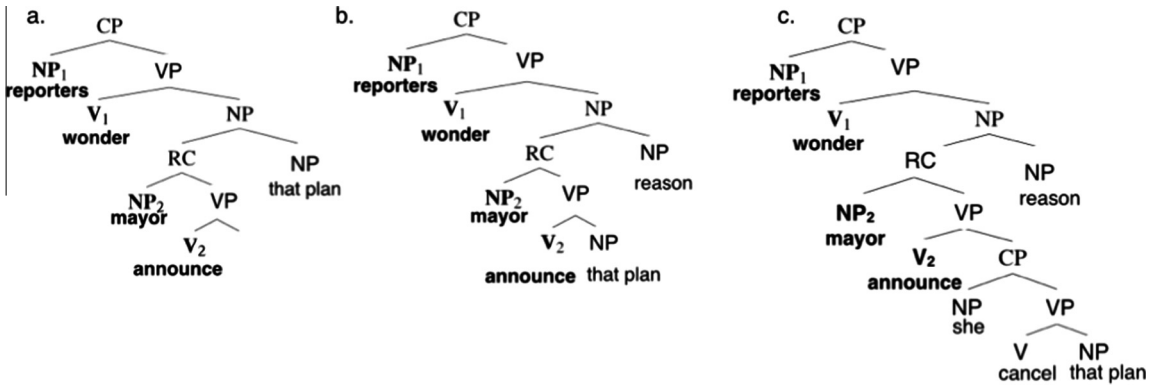
记者们	想知道	市长	宣布	了 那个 计划 的 原因
jizhemen	xiangzhidao	shizhang	xuanbu	le nage jihua DE yuanyin.
reporters	wonder	mayor	announce	perf. that plan DE reason

"The reporters wondered about the reason that the mayor announced that plan."

c.

记者们	想知道	市长	宣布	她 取消了 计划 的 原因
jizhemen	xiangzhidao	shizhang	xuanbu	ta quxiaole nage jihua DE yuanyin
reporters	wonder	mayor	announce	she cancel-perf. that plan DE reason

"The reporters wondered about the reason that the mayor announced that she canceled that plan."



The structures in (10) are all relative clauses. But it is important to note that only (10a) contains an argument gap within the relative clause, i.e. the head noun "the plan" is associated with an argument position within the relative clause. For both (10b) and (10c), the head noun of the relative clause is associated with an adjunct within the relative clause instead of an argument, i.e., the reason of the event depicted by the VP contained within the relative clause. The continuations in (10) are possible alternatives to the right branching wh-dependencies in (9), because Chinese relative clauses are head final. The structures in (9) and (10), therefore, can be continuations from exactly the same initial fragment. The structures in (9) and (10) do not exhaust all possible continuations given the initial fragment in (8), but they could represent two major strategies—either speakers choose to produce a right-branching structure that contains a covert wh-in-situ dependency, or they avoid the right-branching structure and produce an alternative that contains no wh-dependency.

Experiments 1a–1d assess the structural frequency of the right-branching and the alternative relative clause structures by examining the completion rate of these structures under different conditions. The to-be-completed sentence fragments were all taken from the stimuli that will appear in the comprehension experiments (Experiment 2 & 3).

Experiment 1a – completing a N1 V1 N2 V2 fragment

Participants and procedure

Forty native Mandarin speakers participated in this study. They were given sentence fragments that consisted of the first four regions of the target stimuli items that will appear in the comprehension experiments, i.e., "N1 V1 N2

V2" strings, and were instructed to complete the sentences. The preambles were presented visually, and participants completed the sentences in writing. We focused our examination on four conditions: the WH-multi-CP and WH-Serial-Verb conditions, and the Declarative multi-CP and serial-verb conditions (i.e., conditions *a* and *b* in examples (6) and (7)). Sixty-four items were created (see an example in (11)). These items were then distributed into four lists based on a Latin Square design. Each list contained 64 trials and was completed by 10 participants.

(11)

a/b	记者们	***想知道***	市长	***宣布/带领*** _____	WH Multi-CP/ Serial Verb
	jizhemen	xiangzhidao	shizhang	xuanbu/dailing	
	reporter	wonder	mayor	announce/lead	
	NP_1	V_1	NP_2	V_2	
c/d.	记者们	***很高兴***	市长	***宣布/带领*** _____	Declarative Multi-CP/ Serial Verb
	jizhemen	hengaoxing	shizhang	xuanbu/dailing	
	reporter	happy	mayor	announce/lead	
	NP_1	V_1	NP_2	V_2	

Coding and analysis

Before data analysis, trials that contained incomplete or ungrammatical sentences were removed. This affected 4% of the total data. We first coded the range of structures produced under each condition. For both the WH and Declarative conditions, the same five types of structures were identified, which we schematically demonstrate in Fig. 2. The "Right branching clauses" and the "Right branching direct object" (i.e., the A and C types) represent completions similar to the structures in (9); and the "Relative clause object" structure (type B) represents completions similar to the relative clause structures in (10). Since the combination of these three structures accounts for 80% or more of the total completion results, our data analyses below focus on these structures, but interested readers can find more details on the other two types of structures (D and E in Fig. 2) in Appendix A. The coding of the completion results was based on basic structure, but utterances classified in the same group may not share exactly the same structural details. For example, for the Right Branching Clauses structure type (Fig. 2A), the lowest predicate (indicated by "…" under the "CP/FP" projection) was transitive for some completions and was therefore followed by an object, but intransitive for other completions. Or sometimes

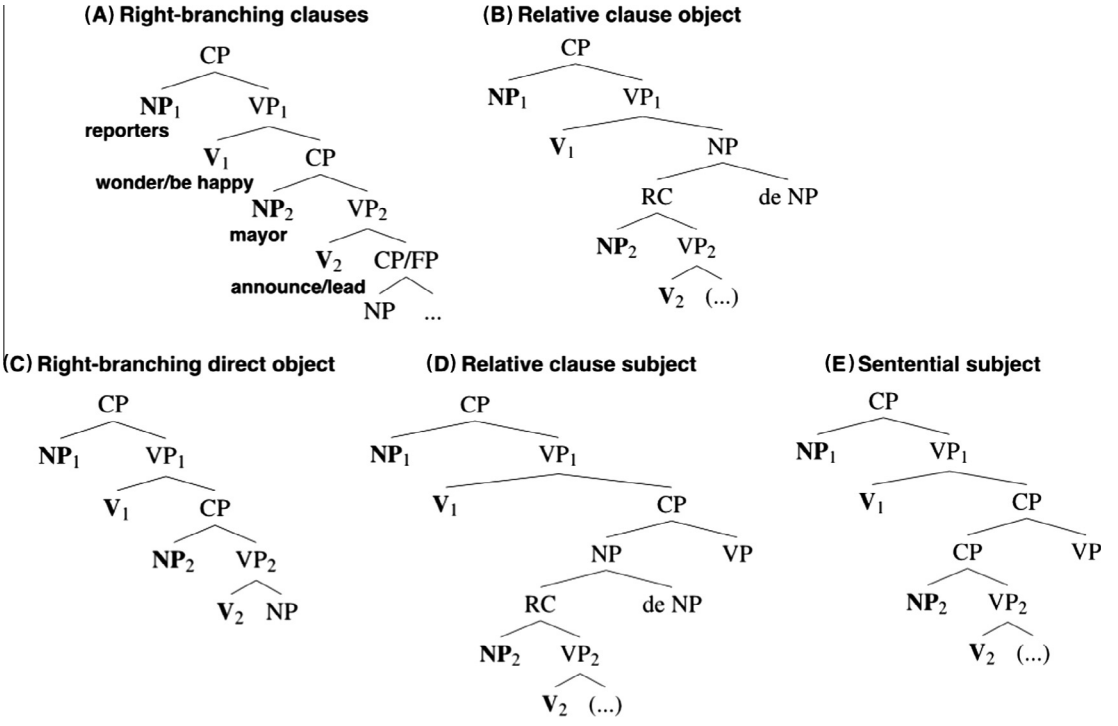


Fig. 2. Top five structures produced in Experiment 1a, when the preamble is an **"N1 V1 N2 V2"** string.

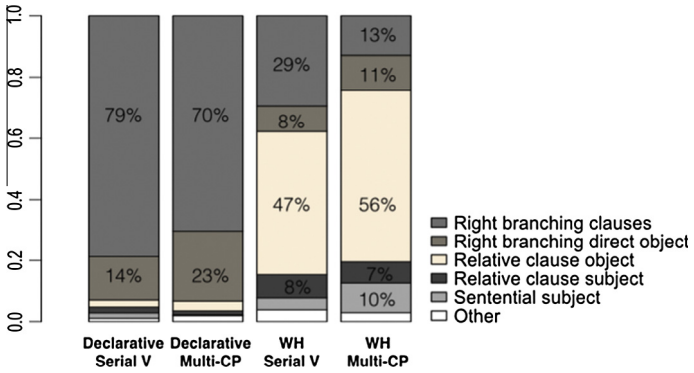


Fig. 3. Experiment 1a, the proportion of each type of structure produced under each condition. Proportions larger than 5% are specified in the plot.

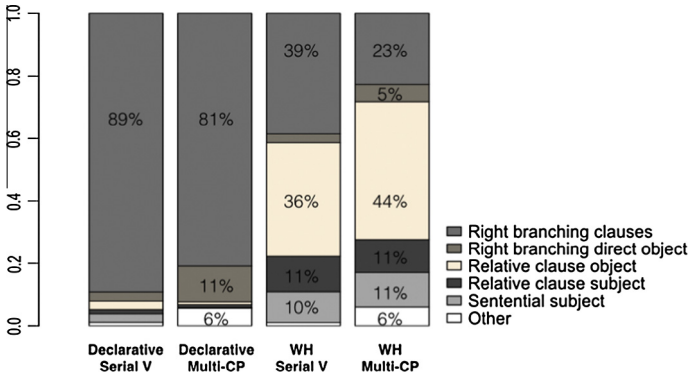


Fig. 4. Experiment 1b, the proportion of each type of structures produced under each condition. Proportions larger than 5% are specified in the plot.

participants chose to include an adverbial phrase for some predicates, but not for others (not shown in Fig. 2).

Results

Distribution of the structures produced. Fig. 3 demonstrates the distribution of the different structures produced under each condition. For the two Declarative conditions, the top two most frequent productions are both right branching structures, including the "right branching clauses" structure (Fig. 8A, 74.5%), in which the continuation itself is a clause and the whole sentence has a right-branching structure (e.g., 12a and 12b below, with the continuation underlined); and the "right branching direct object" structure (Fig. 2C, 18.5%), in which the whole sentence is still a right-branching structure, but the continuation is an NP object (e.g., 12c, continuation underlined).

(12) a.

警员们	听说	警长	承认	自己 贪污
jingyuanmen	tingshuo	jingzhang	chengren	self tanwu
policemen	hear	chief	confess	self embezzle
N1	V1	N2	V2	

"The policemen heard that the chief confessed he embezzled (some money)".

b.

村长	知道	小王	说服	村民 投票
cunzhang	zhidao	xiaowang	shuifu	cunmin toupiao
village-chief	know	Mr. Wang	persuade	people vote
N1	V1	N2	V2	

"The chief of the village knew that Mr. Wang persuaded village members to vote."

c.

球迷们	听说	足联	支持	巴西
qiumimen	tingshuo	zulian	zhichi	baxi
fans	hear	FIFA	support	Brazil
N1	V1	N2	V2	

"The fans heard that FIFA (International Federation of Association Football) supports the Brazil team."

For the WH-conditions, however, the production of right branching structures dropped significantly. Instead, speakers much more frequently produced a relative clause structure (a concealed question more specifically). Some examples are given below in (13). Most of the relative clause structures that participants produced did not contain argument gaps. That is, the head noun—in example (13), *process* or *motivation*—is not associated with any argument gap within the relative clause.

(13)

a.

工人们	想知道	公司	买通	银行 经理 的 经过
gongren-men	xiangzhidao	gongsi	maitong	yinhang jinli DE jingguo
workers	wonder	company	bribe	bank manager DE process
N1	V1	N2	V2	

"The workers wondered about the process that the company bribed the manager of the bank." (i.e., "The workers wondered about the process by which the company bribed the manager of the bank.")

b.

赞助商	想知道	电视台	怂恿	主持人 欺骗 观众 的 目的
canzhushang	xiangzhidao	dianshitai	songyong	zhuchiren qipian guanzhong DE mudi
sponsor	wonder	TV-station	encourage host	lie audience DE motivation
N1	V1	N2	V2	

"The sponsor wondered about the motivation for the TV-station encouraging the show-host to lie to the audience. (i.e., The sponsor wondered about the TV station's motivation for encouraging the host to lie to the audience.)"

For data analysis, we first performed a logistic mixed-effects model, with maximal fixed and random effects, on the production of the "Right Branching Clause" structure. There is a main effect of Structure Type (Declaratives vs. WH, $p < .001$), confirming that the probability of producing the right branching clause structure is higher for Declaratives (74.5%) than for WH-conditions (21%); there is also a main effect of Complexity (Multi-CP vs. Serial Verb, $p < .001$). And crucially, the interaction between Structure Type and Complexity is significant ($p < .05$). For Declaratives, we found no difference between the Multi-CP and Serial Verb condition (70% vs. 79%, $\beta = 0.5$, $SE = 0.34$, $p > .1$). Between the two WH-conditions, on the other hand, there are fewer right branching clause structures under the Multi-CP condition (13% vs. 29%, $\beta = -1.44$, $SE = 0.28$, $p < .001$).

Another logistic mixed-effects model was conducted on the production of the Relative Clause Object structure (structure B in Fig. 2), and we found the expected reversed effect. There were more Relative Clause structures under WH-conditions than under Declaratives (the effect of Structure Type, $p < .0001$). There was an effect of Complexity ($p < .05$), but no interaction between Structure Type and Complexity ($p = .2$). However, planned comparisons suggest that the effect of Complexity was mainly driven by the fact that there were more Relative Clause structures produced under the WH-multi-CP condition than under the WH-Serial-Verb condition (56% vs. 47%, $\beta = 0.53$, $SE = 0.22$, $p < .05$), but there was no difference between the two Declarative conditions (3% vs. 2%, $\beta = 2.2$, $SE = 1.3$, $p > .05$).

Summary of Experiment 1a

Experiment 1a produced two major findings. First, speakers made different syntactic choices under the WH-conditions than under the Declarative-conditions. In particular, many fewer right-branching structures were produced under the WH-conditions. Second, we also found that within WH-conditions, the Multi-CP condition appeared to trigger an "anti-right-branching" bias even more strongly than under the serial-verb condition. Fewer right branching structures were produced under the WH-multi-CP predicate than under the WH-Serial-Verb predicate, and conversely, more relative clause structures were produced under the WH-multi-CP predicates.

Experiment 1b – completing a N1 V1 N2 V2 N3 fragment

Twenty-eight native speakers participated in this study. They were asked to complete sentence fragments that consisted of the first five regions of the target stimuli items, i.e., an "N1 V1 N2 V2 N3" string. The stimuli and the procedure are otherwise identical to those in Experiment 1a.

(14)	记者们	***想知道***	市长	***宣布***	市政府 _____
	jizhemen	xiangzhidao	shizhang	xuanbu	shizhengfu
	reporter	wonder	mayor	announce	city council
	NP_1	V_1	NP_2	V_2	N3

The data analysis procedure was identical to that in Experiment 1a. Ultimately, the distribution of the structures produced under each condition is qualitatively very similar to that in Experiment 1a, as shown in Fig. 4. Again, for the Declaratives, the dominant type of structure participants produced was right branching clauses. For the WH-conditions, on the other hand, the proportion of right branching clause structures is significantly reduced (logistic mixed-effects model: main effect of Structure Type, $p < .0001$), due to the fact that more alternative structures were produced in the WH-conditions, especially relative clause object structures. There was an interaction between Structure Type and Complexity ($p < .05$) when we consider the production of all right branching structures (i.e., pooling together both Right branching clauses and Right branching object structures)—there were fewer right branching structures under the WH-multi-CP condition than under the WH-Serial-Verb condition (28% vs. 41%, $\beta = 0.84$, $SE = 0.3$, $p < .01$), but there was no difference between the two Declarative conditions (92% vs. 91%, $\beta = -0.01$, $SE = 0.37$, $p > .9$). Conversely, more relative clause structures were produced for the WH-multi-CP condition than for the WH-Serial-Verb condition (44% vs. 36%, $\beta = 0.34$, $SE = 0.2$, $p < .1$), whereas there was no difference between the two Declarative conditions (1% vs. 3%).

Experiment 1c – completing a N1 V1 N2 V2 N3 V3 fragment

In Experiment 1c, we created sentence fragments by removing the sentence-final noun phrases from the target stimuli that appeared in the later comprehension experiments, as in the example below. The design was otherwise identical to that of the previous two completion studies. Twenty-eight native speakers participated in this study.

The dominant completion strategy adopted by speakers was to complete (15) with a sentence-final noun phrase, leading to a right-branching structure (Fig. 5A). The mean percentage of this continuation under each condition is: WH Multi-CP: 80.5%; WH Serial verb: 85.9%; Declarative Multi-CP: 99.5%; Declarative Serial verb: 99.8%. The WH-conditions elicited fewer sentence-final-NP continuations than Declaratives (the effect of Structure Type, $p < .05$); this is mainly because, as in the previous two completion studies, WH-conditions elicited more relative clause structures than Declarative conditions (Fig. 5B). Furthermore, between the two WH-conditions, the WH-multi-CP condition was more likely to be continued into a relative clause structure than the WH-Serial-Verb condition (14% vs. 10%, $\beta = 1.94$, $SE = 0.4$, $p < .001$).

A closer look at only the right-branching sentence-final-NP completions under the WH-conditions also reveals another difference between the WH-multi-CP and WH-Serial-Verb conditions. For the WH-multi-CP condition, 68% of the sentence-final NP completions are a Wh-NP (e.g., *which*-NP or *who*), giving rise to an embedded wh-question, as the example in (16); but 32% contain a regular NP followed by either a negation marker ("meiyou") or an interrogative question particle ("ma"), indicating an embedded polar-question (see example (17)). For the WH-Serial-Verb condition, the percentages are 78% and 22% respectively. A mixed-effects logistic model confirms that participants produced significantly more sentence-final Wh-NPs under the WH-Serial-Verb condition than under the WH-multi-CP condition ($\beta = 0.56$, $SE = 0.28$, $p < .05$).

(15)	记者们	***想知道***	市长	***宣布***	市政府	严惩了 _____
	jizhemen	xiangzhidao	shizhang	xuanbu	shizhengfu	yanchengle
	reporter	wonder	mayor	announce	city council	punish
	N_1	V_1	N_2	V_2	N3	V3

(16)

老板	想知道	狗崽队	聘请	侦探	跟踪了	谁
laoban	xiangzhidao	goucaidui	pingqing	zhentan	genzongle	shei
boss	wonder	paparazzi	hire	detective	follow	who
N1	V1	N2	V2	N3	V3	

"The boss wondered who the paparazzi hired the detective to follow."

(17)

小王	想知道	古董店	承认	专家	伪造了	证书 没有/吗.
Xiaowang	xiangzhidao	gudongdian	chengren	zhuanjia	weizaole	zhengshu meiyou/ma.
Mr. Wang	wonder	antique store	admit	expert	forge	certificate not/particle
N1	V1	N2	V2	N3	V3	

"Mr. Wang wondered whether or not the antique store admitted that the expert forged the certificate."

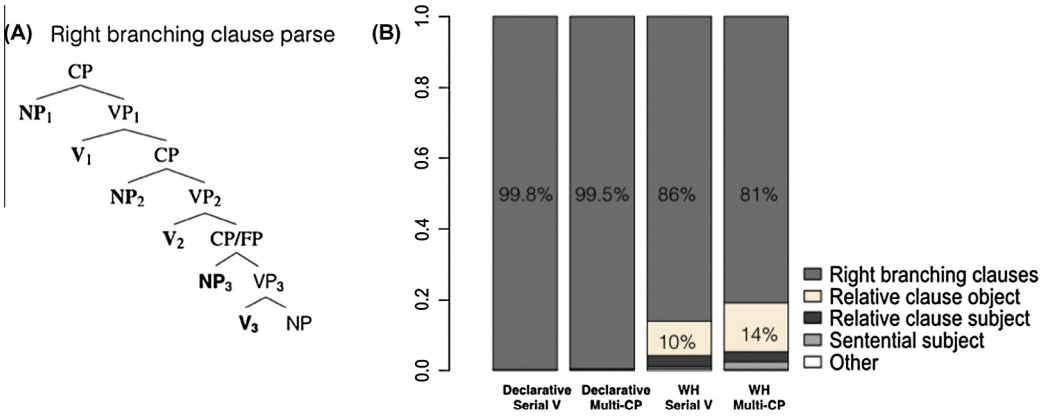


Fig. 5. Experiment 1c. (A) The structure diagram for the sentence-final-NP completion. (B) The proportion of each type of structure produced under each condition. Proportions larger than 5% are specified in the plot.

Since sentence-final NPs, especially the sentence-final Wh-NPs, are the critical regions for the comprehension experiments, in Experiment 1d below we also examined the completion rate of sentence-final Wh-NPs for all WH-conditions.

Experiment 1d—sentence-final Wh-NPs for all WH-conditions

The procedure was identical to Experiment 1c, except that we tested the four WH-conditions (i.e., conditions a–d in example (6)). Thirty-two subjects participated in the study. The predominant continuations given by the participants were sentence-final Wh-NPs, as in (16) above, with 86% for the Short condition, 79% for the Serial Verb condition, 79% for the Adv condition, and 70% for the Multi-CP condition. The Wh-NP completions for the Multi-CP and Serial Verb conditions are overall higher than those in Experiment 1c,[7] but the basic results in Experiment 1c were replicated—there were more sentence-final Wh-NP completions in the Serial Verb condition than in the Multi-CP condition ($\beta = 0.9$, $SE = 0.2$, $p < .01$). Importantly, we also find that when all four WH-conditions were considered, there were also significantly more Wh-NP completions for the WH-Short condition than all other Wh-conditions (all ps < .05).

Summary and discussion of Experiments 1a–1d

The set of completion studies in Experiment 1 generated two major findings. First, there were fewer right-branching structures under WH-conditions than under Declaratives. In other words, seeing a matrix predicate like "wonder" significantly changed speakers' preference for how to continue the given preamble. This observation carries further implications about the production complexity of wh-in-situ dependencies, which we will discuss in the 'General discussion' section.

The second finding is that the presence of a Multi-CP predicate in a WH-condition triggered even fewer right-branching structures than a Serial-Verb predicate. More importantly, both Experiment 1c and 1d revealed that the critical sentence-final Wh-NP occurred less frequently

[7] One possibility is that the sentence-final Wh-NP is an easy and natural continuation for the Short condition, and the inclusion of this condition "primed" the participants to adopt more sentence-final Wh-NPs across the board.

under the Multi-CP predicate than under the Serial-Verb predicate. This raises the possibility that if there is any comprehension complexity difference on the critical Wh-NP between the WH-multi-CP and WH-Serial-Verb conditions, it could either be explained by our original hypothesis that the WH-multi-CP condition is more costly due to memory interference from the intermediate CP structure, or it could be purely due to the structural frequency (or structural expectation) difference between the two conditions.

To distinguish these two possibilities, it is important to look at all four WH-conditions together. As schematically stated in Fig. 1 earlier, the memory retrieval based hypothesis makes the crucial prediction that on the sentence-final-Wh-NP, only the WH-multi-CP condition would trigger interference cost due to the intermediate CP structure; no interference-inspired processing difference is predicted for the other three WH-conditions. The prediction of the frequency/expectation based account, however, is crucially dependent on whether the sentence-final-Wh-NP appears with different frequencies under different WH-conditions. Since Experiment 1d revealed that the completion rate for the sentence-final-Wh-NP is the largest for the WH-Short condition and the smallest for the WH-multi-CP condition, the structural frequency based account would predict the smallest comprehension cost on the critical wh-phrase for the WH-Short condition compared to all other WH-conditions, in addition to predicting the largest cost on the WH-multi-CP condition.

Finally, it should also be noted that even though the structural expectation of a sentence-final-Wh-NP varies under different WH-conditions, overall the sentence-final Wh-NP is the most expected/preferred syntactic outcome given a sentential fragment "N1 V1 N2 V2 N3 V3" (e.g., >70% completion rate in Experiment 1d). This is important since it assures that, in the comprehension experiments below, the critical Wh-NP region is in general unsurprising for the parser.

Experiment 2 – self-paced reading

Participants

Fifty-six native Mandarin Chinese speakers (11 males) between 18 and 26 years old (mean = 21.2) from South China Normal University participated in this experiment.

Material

Sixty-four sets of sentences were constructed for this experiment. Each set contained 8 conditions, based on a 2 × 4 design. The first factor is *Structure Type*: the WH conditions vs. Declarative conditions; and the second factor *Complexity*, with four levels: Multi-CP, Serial Verb, Adv, and Short. Examples are given above in (6) and (7). These 64 sets of experimental items were distributed into 8 lists with a Latin Square design. An additional 35 fillers were also constructed. All of the filler sentences were ungrammatical due to syntactic errors. Half of the fillers contained a wh-word, '哪些' ('which'-classifier), and the other half contained '那些' ('those'-classifier)'. Each subject therefore was tested on 99 sentences total.

Procedure

The experiment was implemented in a self-paced reading paradigm, and stimuli presentation and data collection were carried out using E-prime. Each sentence was presented phrase-by-phrase in the center of the screen. The division of the regions was the same as in the examples shown in (6) and (7). After the presentation of the whole sentence, participants were instructed to rate how acceptable it was on a 1 (totally unacceptable) to 7 (totally acceptable) scale. There were 9 practice sentences before the experimental session.

Data analysis and results

Self-paced RTs

Prior to data analysis, reading times longer than 4000 ms were removed, which affected 2% of the data. Two subjects were removed from the data analysis since more than 40% of their data on the critical word region was lost. For the rest of the 54 subjects, averaged (log-transformed) phrase-by-phrase reading times, starting from the subject of the embedded clause (i.e., the word *mayor* in examples (6) and (7)), are plotted in Fig. 6.

RT means for regions CW-2 to CW are shown in Table 1. For statistical analysis, we performed a mixed-effects linear regression model for each of the 5 regions CW-4 to CW. Raw RTs were first log-transformed before being entered into the models. The models included Structure Type and Complexity as the two fixed effect predictors. Following Barr, Levy, Scheepers, and Tily (2013), we constructed maximal models including all possible random slopes and intercepts for subjects and items, whenever such models converged. Since our models did not reveal any significant effect for the regions CW-4 and CW-3, in our report below we will only focus on the regions CW-2, CW-1, and CW. A summary of the results for the fixed effect terms for all three regions is given in Table 2.

On the CW ("which/that-CL NP")

At the CW region, i.e., the sentence-final noun phrases in (6) and (7), there was a main effect of Complexity (Table 2), suggesting that for both Declarative and WH-conditions, the complexity manipulation modulated the reading times. The critical Complexity × Structure Type interaction is not significant,[8] but planned comparisons within each structure type nonetheless revealed some interesting differences between the WH and the Declarative groups. In Table 3, we present comparisons between the Multi-CP condition and each of the other three conditions, for the WH-conditions and the Declaratives separately, with an absolute $|t| > 2$ indicating a significant difference.

As shown in Table 3, for the Declarative conditions, on the CW (i.e., the sentence-final NP), the Multi-CP condition

[8] We also carried out an analysis with only the Multi-CP and Serial Verb conditions (4 conditions total), the Complexity × Structure interaction is still non-significant ($p = .16$).

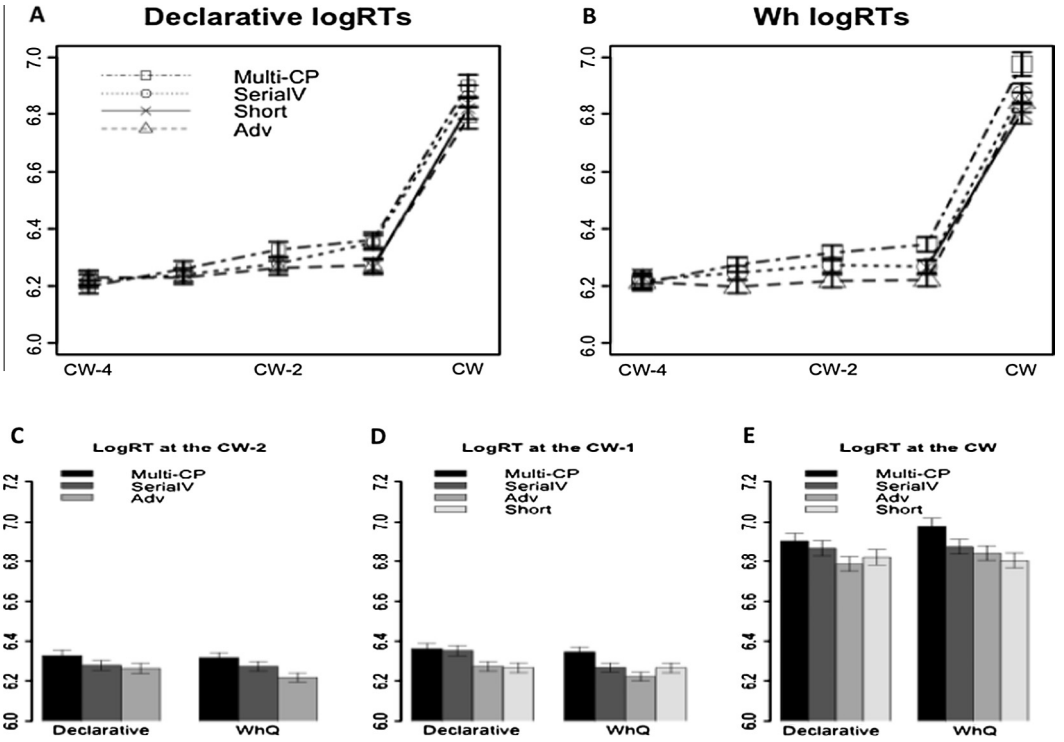


Fig. 6. Experiment 2 phrase-by-phrase reading times. Error bars represent standard errors. Top: Panels (A and B) represent log-transformed RTs from CW-4 to CW, for Declaratives and WH-conditions separately. Bottom: Panels (C–E) plot RTs at each region from CW-2 to CW. Panel (C) (RTs at CW-2) only includes 6 conditions, since the two Short conditions did not have words falling into this region.

Table 1
Experiment 2: Raw RT means (ms) for regions CW-2 to CW. Standard errors are included in parentheses. The "Short" conditions do not have words falling into the CW-2 region (see the description of this region below).

Expt. 2	**CW-2** *("city council")*		**CW-1** *("punished")*		**CW** *("which/that-CL-official")*	
(ms)	WH-conditions Mean	Declarative-conditions Mean	WH-conditions Mean	Declarative-conditions Mean	WH-conditions Mean	Declarative-conditions Mean
Multi-CP	659 (39)	675 (41)	667 (41)	686 (42)	1425 (90)	1296 (79)
Serial Verb	616 (33)	614 (32)	599 (29)	673 (35)	1248 (80)	1254 (78)
Adv	566 (26)	610 (36)	565 (27)	611 (32)	1206 (73)	1154 (73)
Short			608 (32)	611 (31)	1118 (79)	1220 (79)

Table 2
[a]**Experiment 2**: Summary of results on three regions, CW-2 to CW. For CW-2, only 6 conditions were analyzed, excluding the two Short conditions, due to the fact that the Short conditions did not have words falling into the CW-2 region.

Expt. 2	CW-2 *p*-value	CW-1 *p*-value	CW *p*-value
Structure	.3	**<.01****	.20
Complexity	**<.01****	**<.00001*****	**<.00001*****
Structure *Complexity	.6	.10	.30

*$p < .05$.
**$p < .01$.
***$p < .001$.
[a] Whether it is desirable to generate *p*-values for fixed-effect terms in mixed-effects models has been widely discussed recently in the R community. For reporting purposes, we used the Anova (model) in the Car package (Fox & Weisberg, 2011) to generate *p*-values in Tables 2 and 5. A separate set of likelihood ratio tests generated very similar results.

had similar RTs to the Serial verb condition, but it had longer RTs than the other two conditions. Post-hoc pairwise comparisons also revealed that the RTs on the Serial Verb condition are numerically longer than on both the Adv and the Short conditions, but the difference only reached significance for the comparison with the Adv condition (Serial Verb vs. Adv, $p < .05$; Serial Verb vs. Short $p > .1$). It should not come as a surprise that, for the Declaratives, the Multi-CP condition patterned together with the Serial Verb condition, where both elicited a greater processing cost than the other two conditions. As we mentioned earlier, these two long conditions involve complex events (i.e., multiple predicates), which may have led to a greater semantic integration cost on the CW. It is also interesting to note that linear length itself did not necessarily produce more processing complexity. The other long condition—the Adv condition—actually had the fastest

Table 3
Experiment 2: comparisons between the Multi-CP conditions and the other three conditions, for Wh and Declaratives separately, in regions **CW**, **CW-1** and **CW-2**. Within each region, the top row represents the estimate of the Multi-CP condition from a linear mixed-effects model. Each of the other rows represents the difference between other conditions and the Multi-CP condition, with an absolute $|t| > 2$ indicating a significant difference.

	WH-conditions			Declarative-conditions		
	Estimate	Std. error	$\|t\|$-value	Estimate	Std. error	$\|t\|$-value
Expt. 2 CW						
Multi-CP	7.02	0.08		6.95	0.08	
Serial Verb	−**0.13**	**0.04**	**3.08**	−0.05	0.04	1.33
Adv	−**0.16**	**0.04**	**4.10**	−**0.14**	**0.04**	**3.70**
Short	−**0.20**	**0.04**	**4.97**	−**0.11**	**0.04**	**2.78**
CW-1						
Multi-CP	6.35	0.05		6.36	0.05	
Serial Verb	−**0.08**	**0.03**	**2.74**	−0.01	0.03	0.43
Adv	−**0.12**	**0.03**	**4.42**	−**0.09**	**0.03**	**3.23**
Short	−**0.08**	**0.03**	**2.72**	−**0.1**	**0.03**	**3.29**
CW-2						
Multi-CP	6.32	0.05		6.33	0.05	
Serial Verb	−0.05	0.03	1.77	−0.05	0.03	1.66
Adv	−**0.10**	**0.03**	**3.32**	−**0.06**	**0.03**	**2.0**

RTs numerically, suggesting it is the semantic complexity of the event structures, not just linear length alone, that produced the higher integration cost on the Multi-CP and Serial Verb conditions.

For the WH-conditions, our working hypothesis is that the Multi-CP condition should be the most costly since this is the only condition with high interference. This prediction was indeed borne out. In contrast to the Declaratives, the WH-multi-CP condition was read significantly slower than the WH-Serial-Verb condition (Table 3), and also slower than the other two WH-conditions, suggesting that the interference effect led to a processing cost above and beyond any baseline cost associated with semantic integration of complex events. Moreover, the overall results are better captured by the memory retrieval/interference based hypothesis, instead of structural expectation. As discussed earlier, even though both memory interference and structural expectation predict larger processing cost on the WH-multi-CP condition relative to other WH-conditions, the expectation based account, but not the interference account, also predicted the fastest reading time on the WH-Short condition. This was not borne out in the current results. On the CW, we found no significant RT difference between the Short and Serial-Verb WH-conditions ($\beta = 0.07$, $SE = 0.04$, $|t| = 1.6$), or between the Short and Adv Wh-conditions ($\beta = 0.037$, $SE = 0.039$, $|t| = 0.9$).

On the CW-1

CW-1 is the verb before the sentence-final noun phrase, i.e., the verb "*punished*" in examples (6) and (7). An unexpected finding is that the results on this region were very similar to those on the CW. The log-transformed RTs on this region are shown in Fig. 2D. We performed exactly the same analysis as we did on the CW. As shown in Table 2, there was a main effect of Complexity ($p < .0001$), and a main effect of Structure Type ($p < .01$) in this region; and as on the CW, the Complexity × Structure interaction was not significant. In Table 3, we again compared the Multi-CP condition with the other three conditions, for WH and Declarative conditions separately. Our results suggest a very similar trend as for the CW. The WH-multi-CP condition was read significantly slower than the other three WH-conditions (all $|t|$s > 2, $ps < .05$), and pairwise comparisons showed no reliable difference among the other three WH-conditions; whereas for the Declaratives, the Multi-CP condition was not different from the Serial Verb condition, but was significantly slower than the other two conditions, and post hoc comparisons confirmed that the Serial Verb condition had longer RTs than both the Adv and Short conditions ($ps < .01$), while the Adv and the Short conditions were not different ($p > .7$).

On the CW-2

The CW-2 region corresponds to the NP "*city council*" in examples (6) and (7) for the Multi-CP and Serial Verb conditions, and the phrase "*summer*" for the Adv conditions. The two Short conditions did not have words falling in this region. The analysis reported in Table 2 was conducted on all conditions excluding the Short ones. But since the Adv conditions involve different words than the Multi-CP and Serial Verb conditions, we also carried out an additional analysis on only the Multi-CP and Serial Verb conditions (4 conditions total), which again showed a significant effect of Complexity ($p < .05$). We followed up this main effect with pairwise comparisons between the Multi-CP and the Serial Verb conditions, which are shown above in Table 3. For both Declaratives and WH-conditions, there was a trend whereby the Multi-CP condition took longer to read than the Serial Verb condition, although the differences did not reach statistical significance.

We do not have any a priori hypothesis that predicted the observed trend on CW-2. One possibility is that the online expectation of a NP appearing after a Multi-CP predicate is for some reason different from a Serial Verb predicate. Since the NP at the CW-2 region appeared right after a "N1 V1 N2 V2" string, we analyzed the sentence completion results in Experiment 1a to quantify the probability of an overt NP appearing after V2 (NP|V2, the probability of an NP given V2). Our post hoc hypothesis was indeed confirmed: participants produced fewer overt post-V2 NPs when V2 was a Multi-CP predicate, compared to when it was a Serial Verb predicate. The mean percentage of occurrences of an NP after V2 is as follows: WH Multi-CP 58%; WH serial-verb 81%; Declarative Multi-CP 67%; Declarative serial-verb 90%. A logistic mixed-effects model revealed an effect of Complexity (e.g., multi-CP vs. serial verb predicates, $p < .00001$).

Acceptability judgments

For acceptability judgments, there is a main effect of Complexity ($p < .001$), and also a Complexity × Structure Type interaction ($p < .001$). Within Declaratives, the Multi-CP condition was rated lower than the other three conditions ($ps < .001$), and no difference was found among the other three conditions ($ps > .1$). Within the WH-

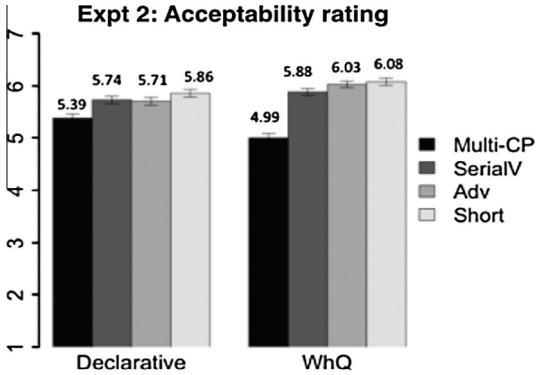


Fig. 7. Acceptability rating for **Experiment 2** on a 1–7 scale, with 7 being the most acceptable.

conditions, the Multi-CP condition is also the one rated the lowest compared to other three conditions ($ps < .0001$); the Serial Verb condition was rated slightly lower than the Adv and Short conditions, but the difference did not reach significance. The crucial interaction between Complexity and Structure Type is driven by the fact that between the two Multi-CP conditions, the WH-multi-CP condition was rated even lower than the Declarative one ($p < .0001$) (see Fig. 7).

Discussion of Experiment 2

To summarize, the most important result revealed by Experiment 2 is that the WH-multi-CP condition evoked extra cost compared to the other WH-conditions, in both online RTs and offline acceptability judgments, confirming our hypothesis that the intermediate CP position induced an additional processing cost. This finding, however, was clouded by two other observations. First, the Complexity × Structure interaction was not significant on the CW; and second, the additional cost associated with the WH-multi-CP condition appeared earlier than we had expected, on the region prior to the sentence final Wh-NP phrase (i.e. at CW-1). We had only expected this cost to emerge on the sentence final WH-phrase, since that is the region where comprehenders received unambiguous information that they needed to initiate a backward retrieval of a higher clause boundary position.

To assess whether the unexpected effect at CW-1 was driven by any online structural expectation factors, we quantified the probability of an overt verb phrase appearing after N3 given a "N1 V1 N2 V2 N3" preamble, using the completion data from Experiment 1b. For each condition, the mean percentage of an overt verb after NP3 is the following: WH Multi-CP 62%; WH serial-verb 62%; Declarative Multi-CP 81%; Declarative serial-verb 89%. There is an effect of Structure Type (mixed-effects model, $p < .0001$), reflecting the observation that in general fewer verb phrases appeared after NP3 in WH-constructions than in Declaratives. But crucially, the probability of a verb phrase appearing after the N3 is the same for the Multi-CP and Serial verb predicates (no effect of Complexity, $p > .8$; no interaction between Complexity and Structure Type, $p > .4$), suggesting online expectation is not a major source of the unexpected effect at CW-1.

It is possible that the unexpected effect on CW-1 is an unreliable observation. It is therefore desirable to assess whether this effect could be replicated. We address this question in Experiment 3 below. More importantly, Experiment 3 will also help us to address a few critical issues. First, Experiment 3 adopted the eyetracking reading paradigm, which presents the stimuli in an ecologically more natural way, and provides more fine-grained timing measures than the self-paced reading paradigm. Second, one concern with Experiment 2 is that the critical Wh-NP region was presented at the sentence-final position, introducing potential noise from wrap-up effects. In Experiment 3, the data analysis will be carried out for the WH-morpheme and the sentence-final NP separately (see more details below), making it possible to more precisely characterize the immediate effect on the WH-morpheme. Finally, given the improvement of the data collection and data analysis methods, Experiment 3 affords us a better opportunity to assess whether the predicted Structure × Complexity interaction is indeed present, in addition to replicating the basic findings in Experiment 2.

Experiment 3 – eyetracking

Material, procedure and participants

The materials used in Experiment 3 were the same as in Experiment 2. Fifty native speakers (34 females, mean age 23 years old) participated in the study. Participants were tested in a sound-attenuated booth. An SR Eyelink 1000 eyetracking system was used to track eye movements at a rate of 1000 Hz. Movements of the right eye were recorded, although viewing was binocular. A Dell SVGA monitor was used to display the stimuli. All stimuli were presented in black against a white background on the monitor. The font Song 22 was used and each character subtended approximately 0.7 degrees of visual angle. Prior to the experimental session, each participant performed a standard 9-point calibration procedure. Before reading each sentence, participants were first asked to fixate on a cross in the center of the computer. A small box then appeared on the left side of the screen where the first word of the sentence would appear. Once the participant correctly fixated on the box, the whole sentence would appear on the screen. After each sentence, participants were instructed to make a binary yes–no acceptability judgment on the sentence they had just read.

Data analysis and results

Acceptability

The main findings from the acceptability result in Experiment 2 were replicated in the binary judgment task in Experiment 3. A mixed-effects logistic model showed a main effect of Complexity ($p < .001$) and an interaction between Complexity and Structure Type ($p < .01$). For both WH-conditions and Declaratives, the Multi-CP condition was accepted the least ($ps < .01$). Furthermore, the

Table 4
Experiment 3: mean RTs (ms) for regions CW-2, CW-1, CW and CW+1, under each condition and each eye movement measure. Standard errors are included in the parenthesis.

Experiment 3	**CW-2** *("city council")*		**CW-1** *("punished")*		**CW** *("which/that-CL")*		**CW+1** *("official")*	
	WH-conditions Mean	Declarative-conditions Mean	WH-conditions Mean	Declarative-conditions Mean	WH-conditions Mean	Declarative-conditions Mean	WH-conditions Mean	Declarative-conditions Mean
FP								
Multi-CP	358 (27)	352 (26)	371 (30)	372 (28)	378 (28)	380 (29)	255 (21)	273 (24)
Serial Verb	327 (22)	333 (24)	372 (28)	371 (28)	378 (29)	382 (27)	285 (20)	289 (24)
Adv			366 (28)	375 (31)	386 (28)	387 (29)	282 (26)	299 (25)
Short			387 (34)	397 (32)	391 (28)	386 (29)	310 (31)	283 (25)
RP								
Multi-CP	479 (61)	485 (59)	611 (113)	665 (123)	1886 (288)	1875 (275)	2109 (317)	2242 (293)
Serial Verb	471 (98)	439 (53)	524 (76)	518 (74)	1151 (159)	1478 (194)	1708 (247)	1675 (230)
Adv			475 (73)	528 (77)	1203 (194)	1377 (174)	1321 (174)	1625 (208)
Short			475 (52)	513 (67)	1025 (135)	1073 (131)	1209 (144)	1222 (154)
TT								
Multi-CP	729 (78)	763 (74)	836 (76)	857 (81)	620 (50)	590 (43)	312 (31)	360 (35)
Serial Verb	555 (52)	559 (56)	684 (62)	723 (65)	559 (48)	607 (51)	346 (28)	346 (30)
Adv			623 (57)	680 (60)	560 (49)	602 (49)	330 (36)	356 (32)
Short			650 (60)	724 (63)	566 (47)	582 (45)	372 (36)	351 (32)

difference between the WH-multi-CP condition and other WH-conditions is larger than the difference between the Declarative-Multi-CP conditions and other Declarative conditions, mainly due to the lower acceptance of the WH-multi-CP condition compared to the declarative Multi-CP condition ($p < .05$).

Eyetracking reading time results

For the RT data analysis, a crucial methodological difference between Experiment 3 and Experiment 2 is how we defined the critical region. For Experiment 3, we split the whole sentence-final noun phrase into two regions. The critical region in the WH-conditions is defined as the WH-morpheme plus the following classifier (two characters total), excluding the head noun. Similarly, for the Declarative conditions, the demonstrative and the following classifier form the critical region, again excluding the head noun. As an example, in (6) and (7), the critical region only encompasses the phrase *which-CL* ('naxie', '哪些') and *those-CL* ('naxie', '那些'), but not the head noun *official* ('guanyuan', '官员'). The newly defined critical region was no longer at the sentence final position, taking care of the methodological concern in Experiment 2.[9] The critical region in the WH-conditions was still the first region that unambiguously signals the presence of a WH-dependency, and therefore all of our experimental predictions remain unchanged. In fact, if our experimental predictions are borne out with this smaller critical region, it provides even stronger evidence that comprehenders immediately start the process of constructing a covert dependency as soon as they reach the WH-morpheme (*which*), without any information about the actual head noun. The regions before the critical region, CW-1, CW-2, etc., are defined in the same way as in Experiment 2 (see (6) and (7)).

Before data analysis we excluded data from any single fixation that was longer than 800 ms or shorter than 80 ms. For the report below, we focus on three measures: First Pass duration time (FP), Regression Path time (RP), and Total Reading time (TT). FP is calculated by summing up all fixations within a region from the first time the eye enters the region from the left until the time the eye leaves, either to the left or to the right. RP includes the first pass duration, plus any subsequent fixations that the eye makes to the left of the region (when reading previous regions of a sentence), until the eye exits the current region for the first time from the right. Both FP and RP measure processing complexity during the early stages of processing. Total reading time (TT), on the other hand, estimates the overall processing complexity at a relatively late stage

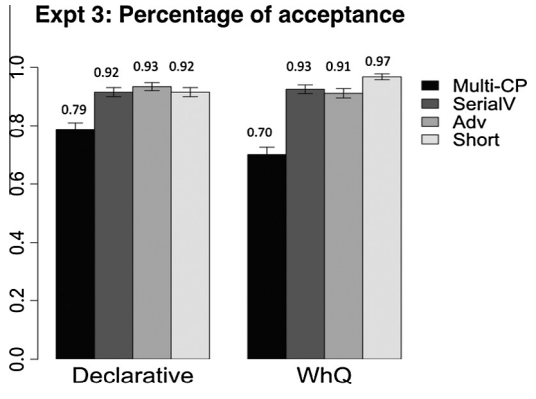


Fig. 8. Yes–no binary acceptability rating for **Experiment 3**.

[9] One anonymous reviewer pointed out that the sentence-final wrap-up effect may not be completely removed from the WH-morpheme. This is because the WH-morpheme is close to the sentence-final position, and the processing of it may (partially) reflect the parafoveal processing of the CW +1. We acknowledge this is indeed a possibility. But given that the CW in different conditions occupies the same sentential position with respect to the sentence final region, if there is any sentence-final wrap-up effect, it would be equal for all conditions. We therefore think the residual sentence-final wrap up effect, if any, is unlikely to have significantly affected our predictions/conclusions.

Table 5
Experiment 3: Summary of the eyetracking results on CW-2, CW-1, CW, and CW+1, under three different eye movement measures.

Experiment 3	CW-2 *p*-value	CW-1 *p*-value	CW *p*-value	CW+1 *p*-value
First pass duration				
Structure	.9	.8	.8	.9
Complexity	**<.05***	.3	.6	<.06$^{\wedge}$
Structure*Complexity	.5	.9	.9	.7
Regression path time				
Structure	.9	.1	**<.01****	.3
Complexity	<.08$^{\wedge}$	.16	**<.00001*****	**<.00001*****
Structure*Complexity	.9	.7	**<.05***	.4
Total time				
Structure	.6	**<.01****	**<.05***	.3
Complexity	**<.00001*****	**<.00001*****	.4	.2
Structure*Complexity	.2	.2	.2	.4

$^{*}p < .05$.
$^{**}p < .01$.
$^{***}p < .001$.
$^{\wedge}p < .1$.

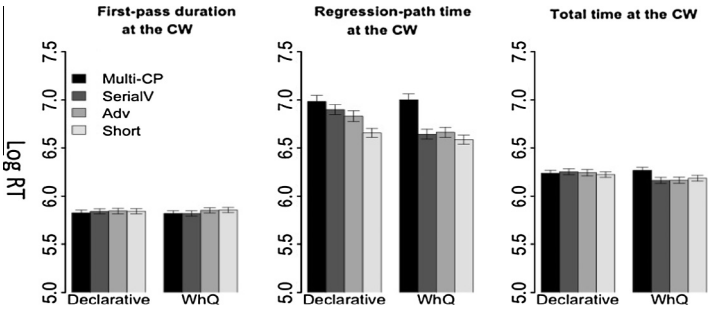


Fig. 9. Experiment 3, three eyetracking measures on the **CW**.

of processing, by summing up all fixations within a region, regardless of whether the eye has exited from this region from the left or the right boundary before.

We report analyses for four regions below: first we analyzed the critical region CW to see if the predicted differences were borne out; second we analyzed the two regions before the CW (i.e., CW-1 and CW-2) to see if there was any pre-existing difference that would impact our interpretation of the result at the CW; and finally we examined the region after the CW, i.e., the sentence-final noun (excluding the WH-morpheme). The raw RT means (ms) for these four regions, for each condition and each eye movement measure, are shown in Table 4. For data analysis, mixed-effects linear regression models were carried out for each of these measures. The modeling procedure was the same as for the analysis of the self-paced reading times in Experiment 2. All data were log-transformed before being entered into the model. A summary of the fixed effects for each region under each measure is presented in Table 5.

On the CW. We did not find any main effect or interaction on the FP measure (all *p*s > .5). We discuss regression path time and total reading time below.

Regression path time (RP). For RP time, there was a main effect of Complexity on the CW ($p < .001$; Table 5), reflecting the observation that for both WH-conditions and Declaratives, the Multi-CP condition had the longest regression path time compared to other conditions (Fig. 9). Crucially, however, this difference is more pronounced in the WH-conditions than in the Declaratives (Complexity × Structure interaction, $p < .05$, Table 5). The critical interaction effect was also significant when we ran the analysis on only the Multi-CP and Serial-Verb conditions (4 conditions total, $p < .05$). To see this interaction effect more clearly, we carried out comparisons between the Multi-CP condition and other conditions, as shown in Table 6. For the WH-conditions, the Multi-CP condition was read significantly slower than the other three conditions; but for Declaratives, the Multi-CP condition was not different from the Serial Verb condition, while it was read more slowly than the Adv and the Short conditions. The crucial interaction assures that there is additional processing complexity only for the WH-multi-CP condition, which goes above and beyond any baseline complexity differences among conditions.

As in Experiment 2, the WH-Short condition was also compared to the WH-Serial-Verb and the WH-Adv conditions. Neither comparison yielded a significant difference (WH-Short vs. Wh-Serial-Verb: $\beta = 0.072$, $SE = 0.07$, $|t| = 1$; WH-Short vs. WH-Adv: $\beta = 0.06$, $SE = 0.07$, $|t| = 0.8$).

Table 6
Experiment 3, on the **CW**. Comparisons between the Multi-CP conditions and the other three conditions, for three different eyetracking measures. Under each measure, the top row represents the estimate of the Multi-CP condition (log-transformed RT) from a linear mixed-effects model, and each of the other rows represents the difference between other conditions and the Multi-CP condition. Absolute |*t*|-values larger than 2 indicate significant differences.

Expt. 3 CW	WH-conditions			Declarative-conditions		
	Estimate	Std. error	|*t*|-value	Estimate	Std. error	|*t*|-value
First pass (FP)						
Multi-CP	5.81	0.04		5.8	0.04	
Serial Verb	0.006	0.04	0.17	0.02	0.04	0.5
Adv	0.03	0.04	0.78	0.03	0.04	0.7
Short	0.04	0.04	1.1	0.03	0.04	0.6
Regression path (RP)						
Multi-CP	7	0.08		7	0.09	
Serial Verb	**−0.34**	**0.08**	**4.48**	−0.1	0.08	1.26
Adv	**−0.35**	**0.08**	**4.39**	**0.17**	**0.08**	**2.07**
Short	**−0.42**	**0.07**	**5.35**	**0.34**	**0.08**	**4.38**
Total time (TT)						
Multi-CP	6.26	0.04		6.23	0.04	
Serial Verb	**−0.1**	**0.04**	**2.28**	0.02	0.04	0.38
Adv	**−0.1**	**0.04**	**2.43**	0.0008	0.04	0.02
Short	−0.08	0.05	1.68	−0.005	0.05	0.11

Total time. Consistent with the results from the Regression Path time measure, the WH-multi-CP condition evoked a longer total time than the other WH-conditions (Table 6, and Fig. 9), but the Declarative-Multi-CP condition was not different from the other Declarative conditions. The Complexity × Structure interaction, however, is not significant in the overall regression model (Table 5, $p = .2$). But a separate mixed-effects model on just the Multi-CP and Serial-Verb conditions (4 conditions total) indeed found a significant Complexity × Structure interaction ($p < .05$).

On the regions before the CW: CW-2 and CW-1. We examined CW-2 and CW-1 to see whether the effect observed on the CW had already emerged in the previous regions. In examples (6) and (7), the region CW-1 is defined as the verb before the sentence final noun phrase, i.e., the verb "punished", and CW-2 is defined as the subject NP of the lowest embedded clause, i.e., the noun phrase "*city council*". Neither of these regions showed the critical differences found on the CW.

On the CW-2, since only the Multi-CP and the Serial Verb conditions had words falling into this region, we did analyses only on these conditions (4 conditions total). As shown in Fig. 10, for both WH-conditions and Declaratives, across all three eye movement measures the Multi-CP condition was more costly in this region than the Serial Verb condition (Table 5, the effect of Complexity, FP time $p < .05$; RP time $p < .1$; TT time $p < .0001$), replicating the self-paced reading results in Experiment 2 on the same region.

The results from CW-1 are also plotted in Fig. 10. Critically different from Experiment 2, effects on CW-1 were not similar to those on the CW. At this region, as shown by the summary in Table 5, neither main effects nor interactions were found for the first pass duration time or the regression path time. The total reading time on CW-1 revealed an effect of Complexity ($p < .0001$) and an effect of Structure type ($p < .01$), but no interaction ($p > .6$). Follow-up analyses for the WH-conditions and the Declaratives separately showed that for both WH-conditions and Declaratives, the Multi-CP condition had longer total reading time than other conditions (|*t*|s > 2; Fig. 10).

On the region after the CW: CW+1. The CW+1 region is defined as the sentence-final NP, excluding the WH-

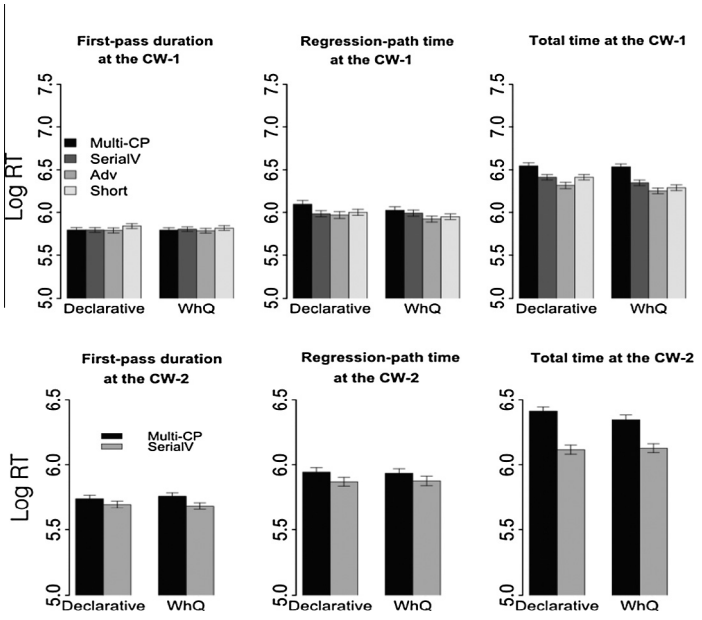


Fig. 10. Experiment 3. Top: results at the **CW-1**; bottom: results at the **CW-2**.

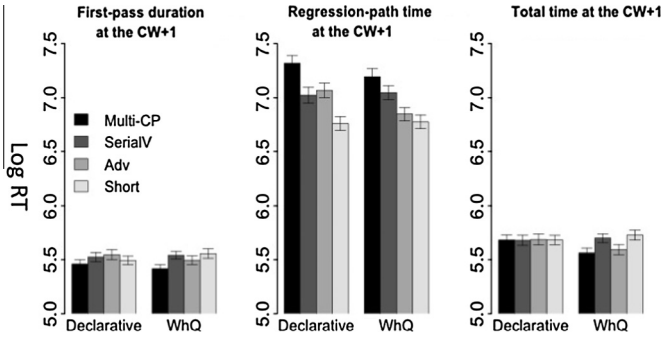


Fig. 11. Experiment 3, three eyetracking measures on the **CW+1**.

morpheme. The results are presented in Fig. 11. In this region, overall, we see very similar effects among the WH-conditions and the Declaratives (see Table 5). There is a near-significant effect of Complexity in the first pass durations ($p < .06$), driven by the fact that for both WH-conditions and Declaratives alike, the Multi-CP condition overall had a shorter first pass duration times than the other conditions. There is also an effect of Complexity in the regression path reading times ($p < .0001$), due to the fact that the Multi-CP condition had a longer regression path reading time compared to other conditions. But there is no interaction between Structure and Complexity for either measure ($ps > .4$).

Discussion of Experiment 3

Using eye tracking, Experiment 3 examined the same set of stimuli as Experiment 2. Methodologically, this experiment presented the stimuli in a more naturalistic setting, and also provided multiple eye movement measures that can better inform us about the processes underlying the comprehension of WH-in-situ questions. We made an additional improvement in the data analysis by splitting the large critical region in Experiment 2—the whole sentence final noun phrase—into two smaller regions: the WH-morpheme part (or the demonstrative morpheme for Declaratives), which is now our critical region, and the head noun. Narrowing down the data analysis to a smaller critical region not only eliminated the potential noise from the sentence-final wrap up effect, but also allowed us to more precisely characterize the immediate effect on the WH-morpheme, lending stronger support to the incremental processing of wh-in-situ dependencies.

Experiment 3 provided evidence that on the WH-morpheme, the WH-multi-CP structure is the most costly among all the WH-conditions, and that this effect goes above and beyond any baseline complexity difference between the Multi-CP construction and other constructions, as indicated by the critical Complexity × Structure Type interaction. The extra processing cost for the WH-multi-CP condition appeared exactly on the critical WH-morpheme; no such effect was found on the pre-critical regions, nor was there any spill-over to the post-critical region. Taking together the comprehension results from Experiments 2 and 3, the overall findings support our original hypothesis: the in-situ wh-phrase triggers the memory retrieval of its associated scope position from a higher clause edge boundary, a process which can be disrupted by interference from intermediate clausal structures. Such an effect strongly indicates the construction of an abstract non-local dependency. Results from processing time were also corroborated by results from acceptability judgments. In both Experiments, the WH-multi-CP condition was rated the lowest among all conditions. It is known that processing complexity can affect speakers' offline acceptability judgments. For instance, average speakers can be very resistant to grammatical sentences containing multiple center-embeddings, due to the almost insurmountable processing difficulty associated with such structures (Chomsky & Miller, 1963). In the current results, we also think the decreased acceptability judgments on the WH-multi-CP condition, similar to the increased online RTs, were driven by the increased processing complexity on this condition due to memory retrieval interference.

General discussion

This paper examined the online processing of Chinese wh-in-situ constructions. The results presented here provide empirical support for the presence of covert dependencies in processing. In the comprehension experiments (Experiments 2 and 3), we showed that wh-in-situ constructions with intervening clausal structures evoke an additional processing cost compared to wh-constructions that are matched in event complexity and/or length but do not contain intermediate embedded clauses. The observed effect cannot be reduced to the difference in baseline complexity between constructions with an embedded clause and those without, as shown by the comparison between wh-conditions and their matched declarative controls. We interpret these findings as the result of memory retrieval interference, lending support to the hypothesis that the parser establishes a covert dependency between the wh-in-situ phrase and a higher scope position. In addition, the set of sentence completion studies in Experiment 1 also revealed that, first, speakers refrained from producing wh-in-situ structures in general; and moreover, the presence of intervening clausal structures further strengthened the bias against the wh-in-situ

structures, and encouraged the speakers to pursue alternative structures. We discuss the implications of both the comprehension and the sentence completion results below.

Memory retrieval in the comprehension of covert dependencies

In the comprehension results, the fact that intervening CP positions can induce interference effects strongly implicates the involvement of a "silent" higher CP position in the construction of a wh-dependency. At the moment at which an in-situ WH-morpheme is encountered, an incremental parser retrieves the relevant higher CP position from working memory. Such a memory retrieval process is subject to the same kind of memory interference that regulates the comprehension of overt non-local dependencies. The only processing difference between overt and covert dependencies is that, in the former case, memory retrieval is often guided by explicit morpho-syntactic cues, whereas in the latter, the parser is guided by structural/configurational properties. Wh-fronting and wh-in-situ languages, therefore, utilize qualitatively very similar processing mechanisms to handle long distance wh-dependencies, despite the typological differences in surface word order.

The interpretation we propose here endorses a core assumption of the cue-based memory retrieval framework that the memory representations probed by the parser are content-addressable—all representations in memory that match one or more of the retrieval cues are directly accessed simultaneously (McElree, 2001; McElree, Foraker, & Dyer, 2003).

There might be alternative interpretations if different assumption are adopted about the memory access mechanism. For instance, with a serial search mechanism, while looking for the scope position for an in-situ wh-phrase, the parser could be implemented to access the most recent clause boundary position first, and if it is not the right target, the parser would move on to the next clause boundary position (see Dillon et al., 2014 for a serial search mechanism in the case of Chinese long distance reflexive binding). This would also predict that wh-constructions with intervening clausal structures are more costly than those without, but without evoking retrieval interference. The current study was not designed to directly examine the theoretical question of how memory representations are accessed, and this would be an important question for future research. But some of our previous work has offered preliminary evidence in favor of a direct access mechanism for wh-in-situ questions. With an SAT paradigm, Xiang, Dillon, Wagers, Liu, and Guo (2014) showed that while longer wh-in-situ questions reduced processing accuracy (as measured by the asymptote of the SAT function), they did not affect processing speed (as measured by the rate and intercept of the SAT function), essentially reflecting a very similar direct access mechanism demonstrated previously for English overt dependencies (e.g., McElree et al., 2003).

Another possible explanation for the current results is that the processing cost on the WH-multi-CP condition is driven by the competition between two potential scope sites, instead of interference from the intermediate [+CP] structure, per se. The lexical properties of a verb like "announce" allow it to optionally take a [+Q] complement. When embedded under the matrix verb "wonder" in the WH-multi-CP condition (e.g., example 6a), since "wonder" obligatorily requires a [+Q] complement, it is ungrammatical to encode "announce" also as [+Q]—i.e., the whole sentence in (6a) is globally unambiguous: the only correct scope position for the wh-in-situ phrase follows "wonder", not "announce" (see footnote 4). However, the possibility remains that the verb "announce" is initially encoded as [+Q, +CP], leading to competition with the correct [+Q, +CP] scope position after "wonder", and therefore additional processing cost in the WH-Multi-CP condition; and the incorrect parse is only eventually eliminated at a later stage of the processing. Based on the current data, we do not have direct evidence to completely rule out this possibility. But for this hypothesis to be true, one has to make an interesting and yet crucial assumption that the parser does not incrementally (or at least not immediately) incorporate contextual constraints (e.g., the presence of "wonder") into the processing of the embedded predicate "announce", allowing "announce" to carry a [+Q] feature until a later processing stage, at which contextual constraints are eventually applied to filter out the ungrammatical parse. We leave the investigation of this hypothesis to future research.

Memory retrieval and syntactic expectation

There is yet another alternative account that could potentially explain the comprehension complexity on the WH-multi-CP condition. As shown in Experiment 1, the WH-multi-CP structure has the lowest completion rate of sentence-final Wh-NP (Experiment 1c, 1d). It is possible that the low frequency of such structures makes the critical word less expected in online comprehension, leading to a greater comprehension cost. This hypothesis would be in line with a class of experience-based models that emphasize how production shapes the statistical regularities of linguistic forms (Gennari & MacDonald, 2008; MacDonald & Christiansen, 2002; Reali & Christianson, 2007) and how parsing complexity is directly predicated upon statistical expectations (Hale, 2001; Levy, 2008). We do believe that expectation is an important predictor of comprehension complexity; however, we also think that it would be too hasty to claim that expectation is the only relevant complexity metric for explaining the current results.

First, as we discussed in Experiment 1, an expectation-based account would make the prediction that the reading time at the critical wh-in-situ phrase should be the fastest for the Short WH-condition among all WH-conditions, due to the fact that the sentence-final wh-phrase is the most expected for the Short WH-condition (Experiment 1d). This prediction was not borne out in either of the two

comprehension experiments. Instead, the reading time results confirmed the predictions of the memory retrieval based account—increased processing cost was only found for the WH-multi-CP condition; but there was no difference between the other WH-conditions that did not contain an intervening clause boundary, which, therefore, would not be targeted by retrieval interference.

Second, it is informative to consider whether syntactic expectation alone could capture the fact that the comprehension difficulty on the wh-in-situ phrase appeared in regression path reading times in Experiment 3, but not first-pass reading times. Admittedly, a clear linking hypothesis between high-level linguistic processing and eye-movement control is not yet completely developed in the parsing literature (Reichle, Rayner, & Pollatsek, 2003; Reichle, Warren, & McConnell, 2009; Staub, 2010). But in an investigation of the comprehension difficulty of object relative clauses, Staub (2010) suggested that under an expectation-based model, backward regression in eye movements—insofar as it is linked to syntactic expectation—is likely to be triggered by a very high surprisal value on the upcoming word, since regression mainly indexes integration failure and the subsequent reanalysis process. Fixation durations, on the other hand, index cost associated with relatively difficult but eventually successful integration of the upcoming word. Staub noted that the surprisal value on the subject of an object relative clause—e.g., *the fireman* in the example *the employee that the fireman noticed*—was estimated to be only slightly lower than the surprisal on the disambiguating verbs of well-known garden path cases—e.g., *fell* in the example *the horse raced passed the barn fell* (Hale, 2001), triggering integration failure and therefore increased regression path time from the RC subject, while leaving the first-pass reading time unaffected. In the current study, the sentence-final Wh-NP, although less expected for the WH-multi-CP condition than other WH-conditions, is overall still the most preferred parse among all alternatives, with a larger than 70% completion rate for all WH-conditions (see Experiment 1d). This is very different from the classic garden-path or object relative clause examples, in which the critical word indicates a strongly dispreferred syntactic parse and therefore induces a high surprisal. In other words, there is no particular reason to expect that the surprisal value on the wh-in-situ morpheme should be large enough to trigger integration failure. The expectation based model is therefore likely to have predicted that the cost on the Multi-CP condition would appear as lengthened first pass fixation durations rather than inflated regression path times alone, contrary to what we found. We do, however, also want to point out that it is possible that the sentential position of our critical region, i.e. the WH-morpheme is close to the sentence-final position, may have contributed to the observation that our effect appeared mostly on the regression path time measure. Our discussion here therefore still remains speculative.

Given these considerations, we think it is unlikely that syntactic expectation alone is driving the comprehension results reported in this study. This is not to claim that comprehenders are not making any syntactic expectations during incremental comprehension of the wh-in-situ construction. Generally speaking, there is good evidence that both expectation-based and memory-based parsing strategies make contributions to online comprehension difficulty (Boston, Hale, Vasishth, & Kliegl, 2011; Levy & Keller, 2013; Staub, 2010). In the particular case we examined here, a matrix predicate like "wonder" strongly signals an interrogative complement, which narrows down the set of possible syntactic forms of the complement. Furthermore, syntactic expectation could be increasingly sharpened with every incoming word. It is yet unclear to us how the expectations stemming from "wonder" are maintained and modulated over time, and how this interacts with the memory retrieval mechanism discussed in this paper. For future investigations, it would be informative to compare the wh-in-situ complements embedded under "wonder" and those embedded under a regular predicate like "know", since the two predicates differ in their cuing strength of the upcoming wh-dependency.

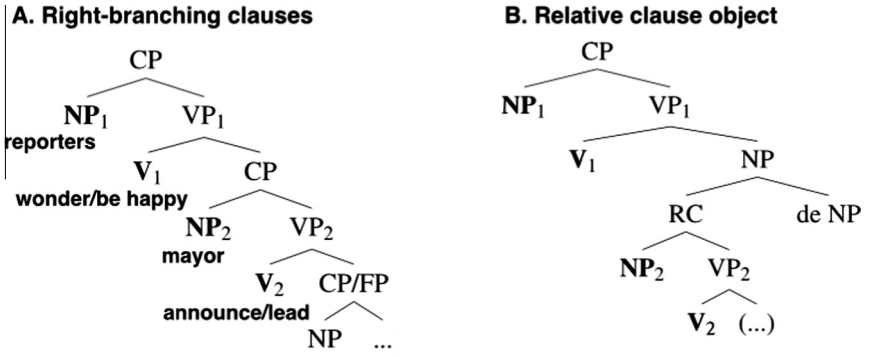
The production complexity of wh-in-situ dependencies—implications from the sentence completion task

The main focus of the current study is the comprehension complexity that emerged from processing wh-in-situ dependencies, but the results from the completion studies in Experiment 1 also have some important implications for the complementary production complexity of wh-in-situ constructions. One major finding is that speakers refrained from producing a default right-branching structure when they encountered a predicate like "wonder", in contrast to a regular predicate like "know" (or "be happy", "hear", etc.). This finding, coupled with some facts about Chinese word order, suggests that wh-in-situ dependencies evoke substantial production complexities, adding empirical support for the parallel between comprehension and production (Pickering & Garrod, 2013).

As discussed earlier in Experiment 1, Chinese grammar allows an initial string "NP1 V1 NP2 V2 ..." to be continued into either a right branching structure or a center-embedded relative clause structure, due to the fact that Chinese relative clauses are head-final. These two structures are depicted in in (18) below (syntactic structures repeated from Fig. 2A and B):

The sentence completion task in Experiment 1 is a hybrid task that taps into both the comprehension/parsing and the production processes. In order to complete the given sentence fragment, participants have to first parse the preamble itself. Given a "NP1 V1 NP2 V2" string in Chinese, under standard assumptions it is a more complex parse to analyze NP2 as starting a center-embedded

(18):



relative clause, rather than a right-branching structure, since the former violates the principle of "minimal attachment" (Frazier & Fodor, 1978)—more syntactic nodes are postulated at NP2 for the relative clause parse. Speakers are indeed sensitive to this complexity difference—they predominantly produced right-branching structures when the matrix verb V1 was a regular predicate like "know" (or *hear*, *discover*, *be happy*, etc.), even though a relative clause structure is also possible under such predicates. However, speakers deviated from their default preference for right branching structures when V1 was a predicate like "wonder". Since such predicates force an interrogative complement of some kind, establishing a right-branching structure would entail a significantly increased likelihood of producing a wh-in-situ dependency. This would imply that, starting at the verb "wonder", the production system needs to plan and maintain a syntactic dependency between the upcoming in-situ wh-phrase and its scope position, i.e., the clause edge position after the verb "wonder". Our results from Experiment 1 suggest that speakers choose to avoid such complexity even at the cost of adopting the otherwise dispreferred alternative—the relative clause structure.

The finding that participants choose to produce a locally more complex structure in service of avoiding a globally more complex structure is reminiscent of previous findings on the production of overt long distance dependencies, especially head initial relative clauses. In a number of different languages (modulo language-specific constraints), such as Italian, Spanish, English, Serbian, and Japanese (Belletti & Contemori, 2009; Gennari, Mirković, & MacDonald, 2012; Montag & MacDonald, 2009; Muñoz & Lago, 2012; Utzeri, 2007), and across different methodologies (sentence completion and picture description tasks) and populations (adults and children), a converging observation is that speakers actively pursue alternative structures in service of avoiding the regular object RC structure. For instance, the passive structure "*the baby who is being held by the woman*" is produced instead of "*the baby who the woman is holding*". The passivization strategy effectively changes an object RC into a subject RC, reducing the length of the dependency produced. The tendency to produce a passive instead of a regular object RC structure is further strengthened when the head noun and the subject of the RC are both animate, as in the example above, in contrast to "*the ball that the woman is holding*" (Gennari et al., 2012; Montag & MacDonald, 2009), suggesting production difficulty due to interference from feature similarity.

The production behavior of object relative clauses falls under a larger class of observations showing that speakers' syntactic choices in production are strategically modulated by the pressure to minimize working memory difficulty, e.g., in situations of heavy NP shift (Gibson, 1998; Hawkins, 1994) or extraposition, which dislocates a relative clause from its head noun (Bader, 2014). The parallel results uncovered here for wh-in-situ dependencies suggest that producing a covert wh-in-situ dependency also requires significant working memory support when the dependency is being planned, maintained and monitored. With a predicate like "wonder", the production system needs to plan for a dependency starting from the clause boundary position after "wonder" and ending at a wh-phrase downstream. The two ends of the dependency, i.e., the scope position after a predicate like "wonder" and the in-situ wh-phrase, need to be closely tracked until the downstream wh-phrase is articulated. In our experiments, such working memory pressure motivated speakers to deviate from producing the default right branching structures, and to produce other alternatives, such as relative clauses. The tendency to produce the alternative relative clause structures is even more pronounced when the planned wh-dependency crosses over an intervening clause boundary. We want to point out that although relative clauses are known to evoke processing cost in production (and comprehension), and at first sight may not be a good alternative to wh-in-situ constructions, the production and comprehension complexity associated with relative

clauses have actually only been attested in the parsing literature on relative clauses with argument gaps, e.g., subject and object relative clauses. The majority of the relative clause structures participants produced in Experiment 1 contained an adjunct gap instead of an argument one (see (10b), (10c) and (13)). It is possible that relative clauses with adjunct gaps may not evoke equivalent cost as those with argument gaps, even though this remains an empirical question that requires further investigation. More importantly, between a right-branching structure that contains a wh-dependency and a center-embedded relative clause structure (that contains an adjunct gap), the former involves a longer dependency since the left hand side of the dependency originates after the matrix verb "wonder" (V1 in 18A), whereas a relative clause dependency would start after the embedded VP (VP2 in 18B). The relative clause structure, therefore, provides a more cost-efficient alternative to the wh-structure, with its reduced production complexity (i.e. a shorter dependency) and a semantic interpretation of a concealed question that satisfies the [+Q] requirement of a predicate like "wonder".

Conclusions

To conclude, our findings provide novel empirical support for covert syntactic representations in real-time processing. In both comprehension and production, we found similar behavioral patterns in the processing of wh-in-situ constructions to those found in the processing of overt dependencies by previous investigations. In particular, comprehension of the wh-in-situ construction is prone to a similarity-based memory interference effect, and the syntactic choices speakers make in production are also modulated by the increased working memory pressure caused by planning and maintaining covert wh-dependencies.

Acknowledgments

We thank Jie Li, Nan Li, WenJia Zhang, JuanhHua Yang, Katina Vradelis, and Caroline Frazia for their assistance in collecting the eyetracking data. We also thank Rachel Kamins and Julian Grove for their proof reading and comments on the manuscript. Preparation of the study and the manuscript was partially supported by grants to SuiPing Wang from the Natural Science Foundation of China [NSF 31271086] and the Pearl River Scholar Funded Scheme in Guandong province [GDUPS 2011], China.

Appendix A

A.1. Completions under the "relative clause subject" type

For this type of completions, the matrix verb V1 is followed by a sentential complement, the subject of which contains a relative clause. A few examples are given below.

(1) Examples from Experiment 1a

a.

警察	说	小王	证实	的 事情 是 假的
jingcha	shuo	Xiaowang	zhengshi	DE shiqing shi jiade
police	say	Mr. Wang	confirm	DE thing is fake
N1	V1	N2	V2	

"The police said the thing that Mr. Wang confirmed was fake." (i.e., "The police said what Mr. Wang confirmed was not truthful.")

b.

媒体	想知道	小王	推荐	的 产品 有 什么 功效
meiti	xiangzhidao	xiaowang	tuijian	DE chanpin you shenme effect
media	wonder	Mr. Wang	recommend	DE product have what effect
N1	V1	N2	V2	

"The media wondered the product that Mr. Wang recommended had what effect." (i.e., "The media wondered about the effect of the product that Mr. Wang recommended.")

c.

媒体	想知道	法官	承认	的 受贿 金额 是 多少
meiti	xiangzhidao	faguan	chengen	DE shouhui jin'e shi duoshao
media	wonder	judge	admit	DE bribe amount is how much
N1	V1	N2	V2	

"The media wondered the embezzlement amount that the judge admitted was how much. (i.e. The media wondered how much money the judge admitted that he received from bribery.")

(2) Examples from Experiment 1b

a.

媒体	说	电影公司	承认	代理律师	说 的 话 都是 事实
meiti	shuo	dianyinggongsi	chengren	daililvshi	shuo DE hua doushi shishi
media	say	film company	admit	attorney	say DE word all fact
N1	V1	N2	V2	N3	

"The media said that the movie company admitted that the words that their attorney said were the facts. (i.e. The media said the movie company confirmed what their attorney said.")

b.

主编	想知道	记者们	协助	小王	完成任务 的 原因 是 什么
zhubian	xiangzhidao	jizhemen	xiezhu	xiaowang	wanchengrenwu DE yuanyin shi shenme
chief-editor	wonder	reporter	help	Mr. Wang	complete task DE reason be what
N1	V1	N2	V2	N3	

"The editor wondered that the reason that the report helped Mr. Wang to finish his task was what. (i.e. The editor wondered why the reporters helped Mr. Wang to complete the task.")

A.2. Completions under the "sentential subject" type

For this type of structures, the matrix verb V1 is also followed by a sentential complement, but the subject of this clause is a clause itself. Some examples are given below.

(3) Examples from Experiment 1a

a.

县长	问助手	村长	鼓励	投资商 投资 是 出于 什么 目的
xianzhang	wen zhushou	cunzhang	guli	touzishang touzi shi chuyu shenme mudi
county-head	ask assistant	village-head	encourage	investor invest is due to what purpose
N1	V1	N3	V2	

"The county head asked (his) assistant why the village head encouraged the investors to invest."

b.

小王	向人打听	工程队	不知道	工程方案 改变 是 谁 的 责任 谁 的 责任
Xiaowang	xiangrendating	gongchengdui	buzhidao	gongchengfangan gaibian shi shei DE zeren
Mr. Wang	ask around	engineering-team	not-know	plan change be who DE responsibility
N1	V1	N2	V2	

"Mr. Wang asked around who was responsible for the fact that the engineering team was not informed about the plan change"

(4) Examples from Experiment 1b

a.

朋友们	说	小王	说服	编辑	离开 报社 是 真的
pengyoumen	shuo	xiaowang	shuifu	bianji	likai baoshe shi zhende
friends	say	Mr. Wang	persuade	editor	leave news-agency be true
N1	V1	N2	V2	N3	

"The friends said that it was true Mr. Wang persuaded the editor to leave the news agency."

b.

警察	想知道	小王	协助	朋友们	犯罪 是 什么 原因
jingcha	xiangzhidao	xiaowang	xiezhu	pengyoumen	fanzui shi shenme yuanyin
police	wonder	Mr. Wang	help	friends	commit-crime be what reason
N1	V1	N2	V2	N3	

"The police wondered why Mr. Wang helped (his) friends to commit a crime."

References

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language, 51*, 23–54.

Aoun, J., & Li, A. Y.-H. (1993). Wh-elements in situ: Syntax or LF? *Linguistic Inquiry, 24*, 199–238.

Bader, M. (2014). Working memory and distance in language production. *Poster presented at architectures and mechanisms for language processing (AmLap), Edinburgh, UK.*

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Belletti, A., & Contemori, C. (2009). Intervention and attraction: On the production of subject and object relatives by Italian (young) children and adults. In *Proceedings of GALA* (pp. 39–52).

Bock, J. K., & Eberhard, K. M. (1993). Meaning, sound, and syntax in English number agreement. *Language and Cognitive Processes, 8*, 57–99.

Bock, J. K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology, 23*, 45–93.

Boston, M. F., Hale, J. T., Vasishth, S., & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes, 26*, 301–349.

Cheng, L. L-S. (1991). *On the typology of wh-questions*. PhD dissertation, MIT.

Cheng, L. L. S. (2003). Wh-in-situ. *Glot International, 7*(4), 103–109.

Chomsky, N., & Miller, G. (1963). Introduction to the formal analysis of natural languages. In R. Robert, R. Bush, D. Luce, & E. Galanter (Eds.). *Handbook of mathematical psychology* (Vol. 2, pp. 269–321). New York: Wiley.

Dillon, B., Chow, W.-Y., Wagers, M., Guo, T., Liu, F., & Phillips, C. (2014). The structure-sensitivity of memory access: evidence from Mandarin Chinese. *Frontiers in Psychology, 5*, 1025. http://dx.doi.org/10.3389/fpsyg.2014.01025.

Ernst, T., & Wang, C. (1995). Object preposing in Mandarin Chinese. *Journal of East Asian Linguistics, 4*, 235–260.

Fox, J., & Weisberg, S. (2011). *An {R} companion to applied regression* (2nd ed.). Thousand Oaks, CA: Sage. <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>.

Frazier, L., & Flores d'Arcais, G. B. (1989). Filler-driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language, 28*, 331–344.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*, 291–325.

Gennari, S. P., Mirković, J., & MacDonald, M. C. (2012). Animacy and competition in relative clause production: A cross-linguistic investigation. *Cognitive Psychology, 65*, 141–176.

Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language, 58*, 161–187.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1–78.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1411–1423.

Gordon, P., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51*, 97–114.

Grano, T. (2012). *Control and restructuring at the syntax-semantics interface*. Ph.D. dissertation, University of Chicago.

Grano, T. (2014). *Control without finiteness contrasts: PRO, aspect, and complementation size in Mandarin Chinese*, ms., Indiana University.

Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics*.

Hale, J. T. (2003). *Grammar, uncertainty and sentence processing*. Ph.D. thesis, The Johns Hopkins University.

Harris, J., Pylkkänen, L., McElree, B., & Frisson, S. (2008). The cost of question concealment: Eye-tracking and MEG evidence. *Brain and Language, 107*, 44–61.

Hawkins, J. (1994). *A performance theory of order and constituency*. Cambridge, UK: Cambridge University Press.

Hu, J., Pan, H.-H., & Xu, L.-J. (2001). Is there a finite vs. nonfinite distinction in Chinese? *Linguistics, 39*, 1117–1148.

Huang, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar*. Ph.D. dissertation, MIT.

Huang, Y. (1994). *The syntax and pragmatics of anaphora: A study with special reference to Chinese*. Cambridge: Cambridge University Press.

Huang, C.-T. J. (1989). Pro-drop in Chinese: A generalized control theory. In O. Jaeggli & K. Safir (Eds.), *The null subject parameter* (pp. 185–214). Dordrecht: Kluwer.

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

Levy, R., & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language, 68*, 199–222.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*, 375–419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science, 10*, 447–454.

Li, Y.-H. A. (1990). *Order and constituency in Mandarin Chinese*. Dordrecht: Kluwer Academic Publishers.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4*, 1–16.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*, 35–54.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 817–835.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*, 67–91.

Miyamoto, E. T., & Takahashi, S. (2002). The processing of wh-phrases and interrogative complementizers in Japanese. In N. Akatuka & S. Strauss (Eds.). *Japanese/Korean linguistics* (Vol. 10, pp. 62–75). Stanford: CSLI.

Montag, J. L., & MacDonald, M. C. (2009). Word order doesn't matter: Relative clause production in English and Japanese. In N. A. Taatgen, & H. van Rijn (Eds.), *Proceedings of the 31th annual conference of the Cognitive Science Society* (pp. 2594–2599).

Muñoz, C., & Lago, S. (2012). Asymmetries in the production of relative clauses: A study in Argentinian Spanish. *Revista Virtual de Estudos da Linguagem (ReVEL), 18*, 64–83.

Nathan, L. (2006). *The interpretation of concealed questions*. PhD thesis, MIT, MA.

Paul, W. (2005a). Low IP area and left periphery in Mandarin Chinese. *Recherches Linguistiques de Vincennes, 33*, 111–133.

Paul, W. (2005b). The *serial verb construction* in Chinese: A tenacious myth and a Gordian knot. *The Linguistic Review, 25*, 367–411.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language, 41*, 427–456.

Pickering, M., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*, 329–392.

Reali, F., & Christianson, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language, 57*, 1–23.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26*, 445–476.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16*, 1–21.

Romero, M. (2006). On concealed questions. In *Proceedings of semantics and linguistic theory XVI* (pp. 208–227). Ithaca, NY: Cornell University.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, 71–86.

Tsai, W.-T. D. (1994). *On economizing the theory of A-bar dependencies*. Ph. D. dissertation, MIT.

Ueno, M., & Kluender, R. (2009). On the processing of Japanese wh-questions: An ERP study. *Brain Research, 1290*, 63–90.

Utzeri, I. (2007). The production and the acquisition of subject and object relative clauses in Italian: A comparative experimental study. *Nanzan Linguistics: Special Issue, 3*(1), 283–313.

Van Dyke, J. A. (2002). *Retrieval effects in sentence parsing and interpretation*. Ph.D. thesis, University of Pittsburgh.

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 407–430.

Van Dyke, J. A., & Johns, C. (2012). Memory interference as a determinant of language comprehension. *Language and Linguistic Compass, 6*(4), 193–211.

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language, 49*, 285–316.

Wagers, M., Lau, E., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language, 61*, 206–237.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition, 85*, 79–112.

Xiang, M., Dillon, B., Wagers, M., Liu, F. Q., & Guo, T. M. (2014). Processing covert dependencies: An SAT study on Mandarin wh-in-situ questions. *Journal of East Asian Linguistics, 23*(2), 207–232.