

How structure-sensitive is the parser? Evidence from Mandarin Chinese

Zhong Chen (Cornell), Lena Jäger, Shravan Vasishth (Potsdam)

1. Introduction

A series of recent articles have revived the discussion about the role grammatical constraints play in online sentence comprehension processes. Phillips, Wagers & Lau (2010) provide a very comprehensive summary of this evidence available in the experimental literature. The conclusion from this body of work is that the human sentence comprehension mechanism (hereafter, the parser) utilizes fairly fine-grained grammatical constraints in real time. A striking example of the parser's use of grammatical information is the finding that the parser does not posit gaps within syntactic islands; this suggests that the parser can use grammatical knowledge about island constraints online when it makes a decision about what structure to build. It is fair to say that the evidence for such a role for grammar is quite well-motivated (although it does not go uncontested, cf. Kluender, 1998, Sag, Hofmeister & Snider, 2007).

A related issue that Phillips et al. (2010) bring up in their review is the search strategy pursued by the parser as it completes dependencies between reflexives and antecedents. Given the existing evidence that the parser utilizes grammatical constraints when making parsing decisions, it is reasonable to assume that this dependency resolution process is informed by grammatical constraints. Here, the relevant grammatical constraint is Principle A of the binding theory. This principle states, approximately, that the antecedent of a reflexive must c-command the reflexive and be in the same clause as the reflexive. If the parser uses Principle A to find the antecedent of a reflexive, then it follows that the search strategy it pursues for completing the dependency between the antecedent and the reflexive will be relatively 'intelligent': the parser only needs to find the relevant noun (this is usually a unique noun, at least in English) that is in the correct syntactic configuration with respect to it. Specifically, when searching the string preceding the reflexive, the parser does not need to consider any candidate other than the correct one.

This is a very attractive idea theoretically: the parser carries out an informed search and efficiently finds the antecedent. Indeed there is apparently considerable evidence consistent with this intelligent search hypothesis.

Evidence in support for this claim comes from Sturt (2003), who ran two eyetracking experiments comparing the effect of gender match of a structurally accessible and a structurally inaccessible noun on the binding of the English reflexive *himself/ herself*. He had a 2x2 factorial design with sentences in which the gender of the reflexive either matched or mismatched the stereotypical gender of the grammatical antecedent as well as the gender of another, structurally inaccessible, proper noun. Sturt reports a main effect of gender match/mismatch of the grammatical antecedent for early measures (first fixation duration, gaze duration, regression path duration) as well as for second pass reading. The gender match/mismatch of the inaccessible antecedent showed no effect in early measures. In second-pass reading time, which is (arguably) considered a late measure, an interaction between the two factors was observed. Sturt concludes that the initial interpretation of the reflexive is driven by Binding Principle A and thus not affected by the presence of structurally inaccessible nouns. He suggests that the late interference effect of an inaccessible antecedent might reflect recovery strategies and wrap-up effects (which in some cases can lead to an incorrect final interpretation of the sentence).

Another important piece of evidence comes from Xiang, Dillon & Phillips (2009), who investigated reflexives in settings like (1) using event-related potentials (they carried out an interesting and important comparison between the processing of reflexives and polarity items, but the present study is only concerned with the processing of reflexives, so we do not discuss the polarity data here).

(1) a. *Congruent*

The tough **soldier** that **Fred** treated in the military hospital introduced **himself** to all the nurses.

b. *Intrusive*

The tough **soldier** that **Katie** treated in the military hospital introduced **herself** to all the nurses.

c. *Incongruent*

The tough **soldier** that *Fred* treated in the military hospital introduced **herself** to all the nurses.

In example (1a), the syntactically licit antecedent for *himself* is *soldier*; *Fred* occurs inside a relative clause modifying *soldier* and cannot therefore be a legitimate antecedent. In (1b) and (1c) the antecedent of the reflexive *herself* is also *soldier* (but the reader would have to reassess at the reflexive the default assumption that *soldier* is male; *soldier* can also be female); the key difference between (1b) and (1c) is that in (1b) a female-referring noun, *Katie*, is present that is not a legal licenser of *herself* but nevertheless matches in gender with *herself*.

In their ERP study, Xiang and colleagues found (inter alia) a non-significant positivity in the 800-1000 ms window in the Intrusive (1b) versus Incongruent (1c) condition ($F(1,27)=2.4$, $p=0.13$); had this effect been statistically significant, it would have been an interference effect. One important point to note here is that, to the extent that it can be interpreted as a non-null result (we return to this issue later), the increased positivity in the intrusive case suggests greater difficulty.

Xiang et al. interpret the above finding as a late interference effect. They also found a marginal centro-anterior negativity in the 250-350 ms interval, but reject it as possibly indicating an early effect of interference because they did not find this effect in the basic comparison of the standard intervals but only in a post hoc analysis driven by visual inspection, and because no previous ERP study has found such an effect in connection with reflexives (Xiang et al., 2009:50). In sum, they argue that if any interference effect does exist in the case of reflexives, it is potentially a late effect. In the initial parsing stages, the parser relies on structural cues to search the antecedent. Xiang et al. also point out that their account (absence of interference effects) may only apply to reflexives, and may not extend to pronouns or logophoric anaphors (2009:52). In other words, structure-sensitive retrieval applies to only one specific case of antecedent resolution.

Based on evidence such as that presented by Sturt (2003) and Xiang et al. (2009), Phillips et al. (2010) propose:

[...] we tentatively suggest that argument reflexives are immune to interference from structurally inaccessible antecedents because antecedents are retrieved using only structural cues.

The above quote applies to configurations such as (1) above. As discussed above, in (1b,c), the only legal antecedent of the reflexive *herself* is *soldier* (note that the fact that soldiers are stereotypically male is irrelevant here because the same amount of difficulty should be experienced in reanalyzing *soldier* as a female in both the interference and no-interference condition). The claim is that the reflexive should never consider the structurally inac-

cessible noun *Katie*, even though it has the feminine gender, just like *herself*. To quote Phillips et al. (2010):

[...] we are suggesting that the person, gender, and number features of reflexives like himself, herself, and themselves play no role in the search for antecedents [...].

The present paper provides initial evidence inconsistent with the above claim.

Our study was motivated by the fact that there is clear evidence in the literature that dependency resolution in parsing is driven by an associative cue-based retrieval process. As an example, consider the work of Van Dyke (2007). In an eyetracking study, Van Dyke compared reading times at a verb (*was complaining*) that was preceded by a string containing one or two grammatical subjects, (2).

(2) a. Non-subject, inanimate-referring intervening noun:

The worker was surprised that the **resident** who was living near the dangerous warehouse **was complaining** about the investigation.

b. Non-subject, human-referring intervening noun:

The worker was surprised that the **resident** who was living near the dangerous **neighbor was complaining** about the investigation.

c. Subject, inanimate-referring intervening noun:

The worker was surprised that the **resident** who said that the warehouse was dangerous **was complaining** about the investigation.

d. Subject, human-referring intervening noun:

The worker was surprised that the **resident** who said that the **neighbor** was dangerous **was complaining** about the investigation.

Here, if, at the moment of dependency completion at the verb, the parser is momentarily confused by the presence of two grammatical subjects, then greater processing difficulty should be seen at the verb in (2c,d) compared to (2a,b). If this is an early process, this difficulty should be seen in early measures. This is in fact what Van Dyke found: in the early measure first-pass reading time, longer fixation durations are observed in (2c,d) (413 and 418 ms, respectively) compared to (2a,b) (376 and 382 ms respectively), all p 's < 0.05. This result is rather puzzling if we assume, following Phillips et al. (2010), that the parser is only sensitive to structural cues and not non-structural cues: after all, here the distractor subject inside the relative clause should never have been considered by the parser if it is using only structural information to find the subject of the verb.

Thus, the structure-sensitive account of parsing would need to have a very limited scope: it would not apply to any configuration other than the reflexive (as mentioned earlier, Phillips et al. 2010 further limit the structure sensitivity of reflexives to those in argument position). Pronouns, logophors and any kind of argument-verb dependency would not engage in structure-sensitive search but only reflexives would.

A simpler theoretical alternative would assume early interference effects in reflexives, just as in other head-dependent resolution processes. We agree with Phillips and colleagues that the parser may be using grammatical constraints to decide on parser actions; but it would be very surprising if the parser were to not use every available piece of information in trying to find the antecedent. In English, one such piece of available information is gender marking: when the reflexive is *himself*, the search process should be informed by the fact that the antecedent does not only have to be c-commanding the reflexive, but must also be masculine. Why would the parser ignore the gender cue to make a decision?

Nevertheless, Phillips and colleagues may be right that the parser ignores gender information in the case of reflexives: after all, c-command provides sufficient information for finding the antecedent. This is what Phillips and colleagues mean when they suggest that gender and other features of reflexives play no role in the search for antecedents: the gender information is redundant because the syntactic constraint already provides all the information needed to find the antecedent. The reason that we feel that use of the gender cue may be unavoidable is that it would require the parser to actively avoid using information when it's available. We see no a priori reason for the parser to parsimoniously give priority to syntactic cues over others; if syntactic constraints were given priority, interference effects such as Van Dyke's (2007) would never be seen.

One potential issue with all previous demonstrations of the absence of an interference effect is that these necessarily depend on null results. Arguing for null results is reasonable when statistical power is relatively high (note: we are not referring to 'observed' power; see Hoenig & Heisey, 2001). However, when statistical power is relatively low, i.e., when the prior probability is low of finding a significant effect when it is in reality present, it is difficult to argue on the basis of null results. This issue has been extensively discussed in statistical theory (Cohen, 1988), but has not received the attention it deserves in psycholinguistics and related areas.

For example, for an interference effect that causes a 30 ms delay in processing, where the standard deviation is 110, for power=0.80, type I

error probability 0.05, the number of participants needed to achieve a significant difference (if there really were one in reality) in a two-sided paired t-test is 108. For psycholinguistic studies, it is quite common to have a sample size of 20, which yields a power of about 0.20. In other words, such an experiment has only a 20% chance of finding an effect that is actually present in nature. As Cohen (1988) has pointed out, one might well ask oneself why we bother to put in so much effort into doing a study where we have a relatively low chance of finding anything at the outset. An effect size of 30 ms is by no means an unusually small number in reading studies. If one finds a significant effect in spite of low statistical power, there are real grounds for drawing a conclusion based on the data; but if one *fails* to find an effect, it is difficult to conclude anything.

It is possible that low statistical power may be masking interference effects in the types of configurations discussed above. An obvious way to demonstrate this point is to carry out an experiment (preferably multiple experiments) involving reflexives where sample size is large enough to give us a reasonable chance of finding an effect if there is one.

Based on the experimental results described below, we suggest that Phillips et al.'s proposal that the parser uses only structural cues to find an antecedent may need to be qualified. We suggest that, although the parser may well use structural cues to find an antecedent, it seems to use more than *only* structural cues. If correct, our findings imply that the parser can be fooled by—i.e., suffer interference effects from—nouns in structurally inaccessible positions that match the reflexive on some search cue other than the structural cue. For example, in the case of (1), the parser may indeed experience interference from the intruding noun *Katie*. The parser's search process may indeed be fallible.

In the remainder of this paper, we present the evidence for fallibility of the parser's search process, and then discuss the implications for theories of sentence comprehension.

2. Experiment

In this experiment we investigate the Mandarin Chinese reflexive *ziji*, which has several interesting properties that make it useful for the study of

cue-based retrieval processes. We introduce the syntax of *ziji* before presenting the details of the experiment.

2.1. The Mandarin Chinese reflexive *ziji*

Ziji is unusual among reflexives in that it can be a long-distance reflexive anaphor that is not restricted in the local clause (Huang, 1984). Furthermore, it is subject to structural as well as pragmatic and semantic constraints (Huang & Liu, 2001). When no antecedent is available in the sentential context, *ziji* generally refers to the speaker (Dillon, Chow, Wagers, Guo, Liu, & Phillips, submitted).

When *ziji* is locally bound, it follows Principle A of Binding Theory (Chomsky, 1981), which means that its antecedent is the subject of the clause it is contained in, as shown in (3a-b). However, violating Binding Principle A, *ziji* can also form long distance dependencies, such as (3c).

- (3) a. Nanhai_i hai-le ziji_i.
 boy_i harm-ASP self_i.
 ‘The boy harmed himself.’
- b. Zhe-pian wenzhang_i shuo nanhai_j hai-le ziji_{i/*i/j}.
 this-CL article_i say boy_j harm-ASP self_{i/*i/j}.
 ‘This article says that the boy harmed himself.’
- c. Nanhai_i shuo zhe-pian wenzhang_j hai-le ziji_{i/*j}.
 boy_i say this-CL article_j harm-ASP self_{i/*j}.
 ‘The boy says that this article harmed him.’

The ability of *ziji* to form long distance dependencies poses a challenge for syntactic theory. In response to this challenge, syntacticians have come up with several explanations aiming to defend Binding Principle A. Cole and Sung (1994) explain the existence of the long distance bound reflexive by assuming that *ziji* crosses clause boundaries via cyclic LF head movement. Huang (1984) proposes that long distance reflexives should be interpreted as a special type of anaphoric pronoun that refers to the subject of the matrix clause. He assumes that there is an underlying representation with the subject of the matrix clause being the ‘speaker’ and the embedded clause that contains *ziji* being a direct quote. The long distance *ziji*, he concludes, should not be considered a reflexive anaphor as defined by Binding Theory. In a later account, Huang claims that the locally bound *ziji* is a common reflexive anaphor following Binding Principle A, whereas the long distance

bound *ziji* is to be explained as a logophor driven by pragmatic constraints. Huang states that this logophoric *ziji* has to be contained in a self-description of its antecedent's referent. This self-description can be a description of a property that the referent explicitly self-ascribes or a description the referent is implicitly disposed to self-ascribe or the one that he self-ascribes via the speaker's perspective (Huang & Liu, 2001; cf. also Pan, 1997).

As mentioned above, there are several constraints on the antecedent of *ziji* that hold for both local and long-distance dependency formations. First, the antecedent has to be a c-commander of *ziji* (Huang & Liu, 2001). (3d) is one example of such cases.

- (3) d. Nanhai_i juede muqin bu zai jia de-shihou
 boy_i think mother not at home time
 jiejie_j yinggai zhaogu hao ziji_{i/j}.
 elder-sister_j should take.care.of well self_{i/j}.

'The boy thinks that his elder sister should take good care of him / herself when their mother is not at home.'

Since *muqin* 'mother' does not c-command *ziji*, it is not a grammatical antecedent of *ziji*. *Nanhai* 'boy' and *jiejie* 'elder sister', however, are c-commanders of *ziji* and thus are both potential antecedents.

As indicated by Huang and Liu (2001), in the case of long distance bound *ziji*, there is one exception to the c-command constraint. If being part of an adjunct clause that precedes the matrix clause, *ziji* can be cataphorically bound to the subject of this matrix clause even though it is not its c-commandee:

- (3) e. Yinwei laoshi ma-le ziji_i nanhai_i henshengqi.
 because teacher scold-ASP self_i boy_i very angry.
 'Because the teacher scolded him, the boy was very angry.'

The second structural constraint on potential antecedents of *ziji* is that *ziji* can only refer to a subject NP (Huang & Liu, 2001).

- (3) f. Nanhai_i song-le nühai_j yi-zhang ziji_{i/*j} hua de tuhua.
 boy_i give-ASP girl_j one-CL self_{i/*j} draw POSS painting.
 'The boy gave the girl a painting that he drew by himself.'

In spite of c-commanding *ziji*, *nühai* 'the girl' is not a grammatical antecedent as it is in the object position of the matrix clause, leaving *nanhai* as the only possible antecedent.

There is one important exception that goes against both the subjectivity as well as the c-command constraint. *Ziji* can refer to a grammatical antecedent that is part of the subject NP (Huang & Liu, 2001).

- (3) g. Nanhai_i de jingyan jiu-le ziji_i.
 boy_i POSS experience save-ASP self_i.
 ‘The boy’s experience saved him.’

In (3g), *nanhai* is the antecedent of *ziji*, although it neither c-commands it, nor is it a subject. Instead, it is a sub-commanding NP modifying the subject NP *jingyan* ‘experience’, which in turn is the head noun of an NP that c-commands *ziji*.

In addition to these structural constraints, *ziji* also exhibits semantic constraints on its antecedent: only animate and sentient referents can build a dependency with *ziji* (Dillon et al., submitted).

- (3) h. Nanhai_i piping-le ziji_i.
 boy_i criticize-ASP self_i.
 ‘The boy criticized himself.’
- i. Nanhai_i shuo zhe-pian wenzhang_j piping-le ziji_{i/*j}.
 boy_i say this-CL article_j criticized self_{i/*j}.
 ‘The boy says that this article criticized him.’
- j. Zhe-pian wenzhang_i shuo nanhai_j piping-le ziji_{i/*j}.
 this-CL article_i say boy_j criticize-ASP self_{i/*j}.
 ‘This article says that the boy criticized himself.’
- k. Nanhai_i shuo jiejie_j piping-le ziji_{i/j}.
 boy_i say elder-sister_j criticize-ASP self_{i/j}.
 ‘The boy says (his) elder sister criticized him / herself.’

(3h-j) each has only one animate subject, which results in unambiguous sentences. (3k), however, has two animate subjects (*nanhai* and *jiejie*), both c-commanding *ziji* and thus candidates for being its antecedent. This leads to a globally ambiguous sentence.

As for the long-distance binding of *ziji*, it is important to note that there are certain cases in which intervening non-antecedents of *ziji* can block this dependency formation. For example, although being in a clear non-antecedent position, the personal pronouns *wo* ‘I’ and *ni* ‘you’ can block a long-distance dependency formation with a third-person NP (3l), a third person NP in the same position, however, does not show any blocking effect (3m) (Huang & Liu, 2001).

- (3) l. Nanhai_i dui wo shuo jiejie_j hen ai ziji_{i*/j}.
 boy_i to I say elder-sister_j very love self_{i*/j}.
 ‘The boy says to me that (his) elder sister loves herself very much.’
- m. Nanhai_i dui muqin shuo jiejie_j hen ai ziji_{i/j}.
 boy_i to mother say elder-sister_j very love self_{i/j}.
 ‘The boy says to (his) mother that (his) elder sister loves him / herself very much.’

Furthermore, Huang and Liu (2001) point out that an interfering local singular NP blocks dependency formation with a long distance plural antecedent (3n), whereas in the reverse case there is no such blocking effect (3o).

- (3) n. Zhe-xie ren_i tingshuo nanhai_j hen ziji_{i*/j}.
 these-CL people_i hear boy_j hate self_{i*/j}.
 ‘These people heard that the boy hated himself.’
- o. Nanhai_i tingshuo zhe-xie ren_j hen ziji_{i/j}.
 boy_i hear these-CL people_j hate self_{i/j}.
 ‘The boy heard that these people hated him / themselves.’

Dillon et al. (submitted) point out that *ziji* is not only a structurally constrained reflexive that can form local as well as long distance bindings, but it is also completely retrospective since there are no cues available indicating the existence of a dependency before encountering *ziji*. They argue that exactly these properties of *ziji* make it a very useful tool to test memory access.

The important properties of *ziji* that we exploit in the next experiment are: the possibility of long- and short-distance antecedents; and the requirement common to both long- and short-distance antecedents that they have to be subjects (modulo the exception discussed above) and animate.

2.2. Participants

120 participants from Dalian and Nanjing took part in this experiment for payment (2 Euros per participant).

2.3. Method

We used the self-paced reading methodology (Just, Carpenter & Wolley, 1982). A session started with practice trials to prepare the participants for

the task. Participants are instructed to read at a natural pace. Each trial begins with a screen presenting a sentence in which the words are masked by dashes. The participant has to press the space bar to reveal the next word. After each sentence, a yes-no comprehension question appears on the screen. Participants have to press a key for 'yes' or 'no' responses.

2.4. Design and Predictions

In Chinese structures like (4), the antecedent ‘the opposition leader’ is the only legal antecedent for the reflexive *ziji*, ‘self’ (which requires an animate antecedent, as discussed above). The non-local antecedent case (4a,b) is interesting because it helps us test another prediction of the cue-based retrieval model: due to decay and/or interference from other chunks in memory, the long-distance antecedent would be harder to retrieve at the reflexive, resulting in a stronger preference for a more locally available partially matching candidate. In other words, the cue-based retrieval account predicts an interaction between locality and interference, which the structural-cue based access account does not (since the interfering noun would never be considered as a candidate in either short- or long-distance antecedent cases).

Thus, in (4), under the exclusively structure-sensitive search view, the parser should never consider an intervening noun like *kangyizhe*, ‘protester’, as an antecedent because it is inside an adverbial phrase and cannot c-command the reflexive *ziji*; this predicts no reading time difference between cases where the intervening noun is ‘protest’ (4a,c) versus ‘protestor’ (4b,d). By contrast, the cue-based retrieval account predicts an interference effect (slower reading time in (4b,d) versus (4a,c)), and an interaction between locality and interference, in that the non-local condition should show a stronger interference effect.

(4) a. Long-distance dependency; inanimate interposed NP

Fanduipai-lingxiu	biaoshi	[zhe-ge	shengming	[zai
opposition.leader	said	this-CL	announcement at	
<i>kangyi</i> shikong	de-shihou] _{AdvP}	gaojie-le	ziji	de
protest out.of.control	time	warn-ASP	ziji	POSS
dangyuan] _{CP}				
party.member				

‘The opposition leader said that this announcement warned his party

members when the protest was out of control.’

b. Long-distance dependency; animate interposed NP

Fanduipai-lingxiu	biaoshi	[zhe-ge	shengming	[zai
opposition.leader	said	this-CL	announcement	at
<i>kangyizhe</i>	shikong	de-shihou] _{AdvP}	gaojie-le	ziji de
protester	out.of.control	time	warn-ASP	ziji POSS
dangyuan] _{CP} .				
party.member				

‘The opposition leader said that this announcement warned his party members when protesters were out of control.’

c. Local dependency; inanimate interposed NP

Zhe-ge	shengming	biaoshi	[fanduipai-lingxiu	[zai
this-CL	announcement	said	opposition leader	at
<i>kangyi</i>	shikong	de-shihou] _{AdvP}	gaojie-le	ziji de
protest	out.of.control	time	warn-ASP	ziji POSS
dangyuan] _{CP} .				
party.member				

‘This announcement said that the opposition leader warned his party members when the protest was out of control.’

d. Local dependency; animate interposed NP

Zhe-ge	shengming	biaoshi	[fanduipai-lingxiu	[zai
this-CL	announcement	said	opposition leader	at
<i>kangyizhe</i>	shikong	de-shihou] _{AdvP}	gaojie-le	ziji de
protester	out.of.control	time	warn-ASP	ziji POSS
dangyuan] _{CP} .				
party.member				

‘This announcement said that the opposition leader warned his party members when protesters were out of control.’

We used twenty-four sets of items shown as in (4); these (along with data and R code used) are available online from the website of the Potsdam Mind Research Repository (<http://www.psych.uni-potsdam.de/pmr2>). In addition to the stimulus items, seventy fillers with varying syntactic structures were randomly interspersed between items, with the constraint that at least one filler intervened between two items. Both items and fillers were presented in simplified Chinese characters. Each target item was followed

by a Yes/No question; the answer to this question required the reader to correctly resolve the antecedent-reflexive relationship. For the examples in (4), the corresponding questions would be as follows:

Qa: Did this announcement warn members of the opposition party? Y

Qb: Did this announcement warn members of the opposition party? Y

Qc: Did the opposition leader warn members of the ruling party? N

Qd: Did the opposition leader warn members of the ruling party? N

The non-local conditions always had the correct answer as Y, and the local conditions always had the correct answer as N. We return to this point in the Results and Discussion sections.

2.5. Results

2.5.1. Question-response accuracies and latencies

We first present the analyses for question response latency and accuracy, and then the reading time results. The question-response latencies and accuracies for the four conditions are presented in Tables 1, 2 below.

Table 1. Means and standard errors of response times (ms) in the four conditions.

	Non-local	Local
Non-interfering	(a) 2787 (69)	(c) 2520 (63)
Interfering	(b) 2944 (71)	(d) 2665 (88)

Table 2. Means and standard errors of response accuracy (percentages) in the four conditions.

	Non-local	Local
Non-interfering	(a) 85.62 (0.02)	(c) 87.08 (0.02)
Interfering	(b) 79.79 (0.02)	(d) 88.54 (0.01)

For the analysis of question-response latencies, a linear mixed model (Bates & Sarkar, 2007) was fit with participants and items as crossed random factors, and locality, interference, and their interaction as orthogonally coded

factors. For question response accuracies, a generalized linear model with a logistic link function was fit (a correct response was marked 1, and an incorrect one 0).

As shown in Tables 3,4, we found a main effect of locality in both latencies and accuracies: the non-local condition had longer response latencies, and lower accuracies. The response latency to the interference conditions was significantly slower, but no significant effect of interference was seen in response accuracy. Finally, a significant interaction between locality and interference was found in response accuracies: the interference effect was larger in the non-local compared to local conditions (this is as predicted by the cue-based retrieval model). In question-response latencies this interaction was not significant.

The results shown for question response latencies have 26 extreme values (>10,000 ms) removed; this constituted 1.35% of the data points. If these data points are included, only the locality effect remains significant. No data were removed for the question response accuracies (these are 0,1 scores).

Table 3. Statistical analysis of the question-response latencies. Latencies greater than 10,000 ms were removed (1.35% of the data)—see text for discussion.

Contrast	Coefficient	SE	t-score	p-value
Locality	-0.11	0.02	-6.10	<0.05
Interference	0.03	0.02	2.14	<0.05
Loc x Int	-0.03	0.02	-1.75	n.s.

Table 4. Statistical analysis of the question-response accuracies.

Contrast	Coefficient	SE	z-score	p-value
Locality	0.49	0.14	3.44	<0.01
Interference	-0.18	0.14	-1.23	n.s.
Loc x Int	0.32	0.14	2.23	<0.05

2.5.2. Reading times at the critical and post-critical regions

We present next the statistical analyses for the critical region *ziji* and the post-critical region (the word *de* immediately following the critical region). The post-critical region is interesting because processing costs that arise due to difficulty in the critical region are often expressed in the following region, so-called spillover (Mitchell, 1984).

We removed all reading times longer than 2000 ms (1.1% of the data) because these extreme values skew the residuals of the linear mixed model. The data analysis was carried out on log-transformed reading times in order to respect the additivity assumption of linear mixed models (Gelman & Hill, 2007).

As summarized in Tables 5 and 6, the results show a marginally significant locality effect at the critical region (the reflexive *ziji*), and a significant interference effect. A marginal interaction is also found. If all the data are retained (i.e., if the extreme values greater than 2000 ms are included), then a significant locality effect is seen ($t=-2.02$), and the interference effect is rendered marginal (1.69).

In the spillover region (the word *de* following the reflexive), a statistically significant effect of locality and interference was seen (here, removal of all reading times greater than 2000 ms resulted in removal of 0.3% of the data; results do not change even if we retain these extreme values). No interaction is found.

Table 5. Means and standard errors of reading times (ms) in the four conditions, at the critical region and the region following the critical region (spillover region). These means and standard errors are computed after removing 1.1% of the data points (reading times greater than 2000 ms).

Region	Long-distance, inanimate	Long-distance, animate	Local, inanimate	Local, animate
<i>ziji</i>	410 (8)	448 (12)	410 (8)	415 (9)
<i>de</i>	371 (7)	387 (7)	363 (5)	372 (7)

Table 6. Statistical data analysis of the critical and spillover regions.

Region	Contrast	Coefficient	SE	t-value
ziji	Locality	-0.03	0.01	-1.92
	Interference	0.03	0.01	2.03*
	Loc x Int	-0.03	0.01	-1.97
de	Locality	-0.02	0.01	-2.26*
	Interference	0.02	0.01	2.25*
	Loc x Int	-0.01	0.01	-0.94

2.6. Discussion

To summarize the results, question-response accuracy was higher and question-response latency was shorter in the local conditions, suggesting that processing the local-antecedent case was easier (this is consistent with the findings of Dillon, et al., submitted; Li & Zhou, 2010). In question-response accuracy and accuracy we found an interaction between the locality and interference manipulation.

Regarding reading times, we found: (i) a marginal effect of locality at the critical region, and a significant effect in the spillover region: the local-antecedent conditions were read faster; (ii) an effect of interference at the reflexive and in the spillover region: the conditions with the interfering noun were read slower; and (iii) a marginal interaction in the reflexives region.

The presence of interference effects in the question-response data and the reading time data is consistent with the possibility that not only structural cues are used in searching for antecedents. The interaction seen in question-response data and the marginal interaction ($t=-1.97$) in the reading time data are consistent with the prediction that interference effects should be stronger in the cases where the correct antecedent has decayed more (the non-local case has a stronger interference effect). Note, however, that the evidence for this interaction in the online data is quite weak.

The evidence for interference effects in the question-response data is interesting because it suggests that in the interference conditions the incorrect antecedent has been retrieved and connected with the reflexive. This is clear evidence that the wrong antecedent has been retrieved and has permanently become associated with the reflexive.

The self-paced reading data may be dismissed on the grounds that SPR data might be indexing only later processes. Xiang et al. (2009) have argued that illicit antecedents are considered only in later stages of processing. This objection can be addressed by using methods (such as EEG, and eyetracking) that can help distinguish between early and late processes. We intend to pursue this issue in future work.

One potential concern in interpreting the question-response accuracies is that the non-local conditions always required “yes” as the correct response, whereas the local condition always required a “no” response. This could have led participants to adopt a strategy that could have biased our results. It is difficult to determine whether this is a true confound in the design, but we intend to revisit this experiment with a more counterbalanced design in the questions.

3. General Discussion

The presence of the interference effect is consistent with the view that retrieval processes involve the use of non-structural retrieval cues such as animacy. The alternative view, that only structural cues would be used in antecedent resolution of reflexives, has in our opinion one major flaw: the results are necessarily based on null results (the absence of interference effects, or the inability to observe interference effects). One possibility worth exploring is that these null results may have been observed in experiments with relatively low statistical power. This is a testable claim: if the experiments supporting the structural-cue based explanation are replicated with a larger sample size (how large the sample size needs to be can be estimated from previous results), it is possible that interference effects would be observed. It remains to be seen whether this will turn out to be correct.

Our findings should not be taken to imply that the parser does not employ structural cues to complete dependencies. Our claim is merely that the parser may not ignore relevant cues such as animacy in searching for an antecedent.

We turn now to potential objections to our findings. A legitimate concern in the present experiment is that we only have one experimental result that supports the interference account (but see Badecker and Straub 2002 for evidence from English consistent with ours), whereas the alternative view has several experiments to back up their claims. This is a valid criti-

cism. There is only one way to respond to this objection: we are in the process of trying to replicate this effect not just for Chinese but also for English and other languages.

A further objection could be that the interference effect seems to occur only in the long-distance reflexive condition; as mentioned above, long-distance anaphora in Chinese is believed to be an instance of logophoric anaphora. Xiang et al. (2009) mention that their constraint does not apply to logophoric anaphora. It is possible that our result is restricted to logophors. An easy way to address this concern is to carry out an experiment with English that exactly replicates the Xiang et al. configuration. If an interference effect is seen in such configurations as well, this would suggest that there is nothing special about interference effects. We hope to definitively answer this question before long, but at least one replication (Patil, Vasishth, Lewis, 2011) using eyetracking of the Xiang et al. 2009 materials shows interference effects in first-pass regression probability, arguably an early effect.

To conclude, we suggest in this paper that the evidence for structure-sensitive search in the case of reflexives may not be well-motivated: it is based on null results that may have been a consequence of low statistical power, and in cases such as Xiang et al. 2009, the patterns are in fact consistent with the presence of an early interference effect. Although they interpret the early statistically non-significant effects as null results, we contend that if the experiment were given a chance to show an effect—by increasing statistical power sufficiently—these effects would be observed in their early time window. A further weakness of the structure-sensitive search account for reflexives is its limited scope: a large number of configurations must be exempted from this structure-sensitive search constraint (pronoun, logophors, and argument-head dependencies), allowing only one narrow category to engage in structure-sensitive search. A simpler account would capture the uniform presence of interference effects in all these structures. Of course, the evidence showing interference effects needs to be strengthened by replication across a variety of structures, languages, and methods. For now, the debate remains open.

Acknowledgements

Our greatest debt is to Brian Dillon, Colin Phillips, and Ming Xiang for patiently helping us understand their positions on the issue. This work has benefitted from

discussions with John Hale, Richard Lewis, Umesh Patil, Titus von der Malsburg and Masaya Yoshida. Thanks go to Professor Baojia Li, Nanjing Normal University for allowing Zhong Chen to conduct half the study there, and Professor Huili Wang, Dalian University of Technology for allowing Lena Jäger to visit and carry out the second half of the data collection there. For questions about this paper, please contact Shравan Vasishth (vasishth@uni-potsdam.de).

References

- Badecker, W. & Straub, K.
2002 The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 748–769.
- Bates, D. & Sarkar, D.
2007 lme4: Linear mixed effects model using S4 classes.
- Chomsky, N.
1981 *Lectures on Government and Binding*. Dordrecht: Foris.
- Cohen, J.
1988 *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Cole, P. & Sung, L-M.
1994 Head movement and long-distance reflexives. *Linguistic Inquiry*, 25, 355–406.
- Dillon, B., Chow, W.Y., Wagers, M., Guo, T., Liu, F., & Phillips, C.
submitted The structure-sensitivity of memory access: evidence from Mandarin Chinese.
- Gelman, A. & Hill, J.
2007 *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- Hoening, J.M. & Heisey, D.M.
2001 The abuse of power. *The American Statistician*, 55(1), 19-24.
- Huang, C-T.J.
1984 On the distribution and reference of empty pronouns. *Linguistic Inquiry*, 15, 531–574.
- Huang, C.-T. J. & Liu, C-S. L.
2001 Logophoricity, attitudes, and ziji at the interface. In: P. Cole, G. Hermon, & C.-T. J. Huang (eds.), *Long-Distance Reflexives*. New York: Academic Press, 141–195.
- Just, M.A., Carpenter, P.A. & Woolley, J.D.,
1982 Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111(2), 222–238.

- Kluender, R.
1998 On the distinction between weak and strong islands: A processing perspective. *Syntax and Semantics*, 29, 241–280.
- Li, X. & Zhou, X.
2010 Who is *ziji*? ERP responses to the Chinese reflexive pronoun during sentence comprehension. *Brain Research*, 1331, 96–104.
- Mitchell, D. C.
1984 An evaluation of subject-paced reading tasks and other methods of investigating immediate processes in reading. In: D. E. Kieras & M.A. Just (eds.), *New Methods in Reading Comprehension Research*. Hillsdale, NJ: Lawrence Erlbaum.
- Pan, H.
1997 *Constraints on reflexivization in Mandarin Chinese*. New York: Garland Publishing Inc.
- Patil, U, Vasishth, S., & Lewis, R.L.
2011 Early retrieval interference in syntax-guided antecedent-search? Talk presented at the 24th CUNY Sentence Processing Conference, Stanford, CA.
- Phillips, C., Wagers, M., & Lau, E.
2010 Grammatical illusions and selective fallibility in real-time language comprehension. To appear in J. Runner (ed.), *Experiments at the Interfaces, Syntax and Semantics*, 37. Bingley, UK: Emerald Publications.
- Sag, I., Hofmeister, P. & Snider, N.
2007 Processing complexity in subadjacency violations: The complex noun phrase constraint. *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, 43(1), 215-229.
- Sturt, P.,
2003 The time course of the application of binding constraints. *Journal of Memory and Language*, 48, 542-562.
- Van Dyke, J. A.
2007 Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(2), 407-430.
- Xiang, M., Dillon, B., & Phillips, C.
2009 Illusory licensing effects across dependency types: ERP evidence. *Brain and Language*, 108(1), 40-55.