Language, Cognition and Neuroscience, 2014
Vol. 29, No. 10, 1205–1226, http://dx.doi.org/10.1080/01690965.2013.841969

Routledge
Taylor & Francis Group

Building Chinese relative clause structures with lexical and syntactic cues: evidence from visual world eye-tracking and reading times

Fuyun Wu[a]*, Yingyi Luo[b] and Xiaolin Zhou[b,c,d]*

[a]*Institute of Linguistic Studies, Shanghai International Studies University, Shanghai 200083, PR China;* [b]*Center for Brain and Cognitive Sciences, Psychology Department, Peking University, Beijing 100871, PR China;* [c]*Key Laboratory of Machine Perception and Key Laboratory of Computational Linguistics (Ministry of Education), Peking University, Beijing 100871, PR China;* [d]*PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, PR China*

(*Received 12 December 2011; accepted 18 July 2013*)

Relative clauses (RCs) in Chinese are prenominal. In object-modifying, object-extracted RCs (e.g. *Click on [$_{RC}$ the ball broke] window*, meaning 'Click on the window [$_{RC}$ that the ball broke]), the ambiguous status of the local noun *ball* and the long-distance attachment of the head noun *window* into the main verb appear to make online parsing of Chinese RCs particularly difficult. By interposing mismatching classifiers and the passive marker BEI into the RC sentences, we investigated whether the presence of incomplete heads would add storage costs, as predicted by the Dependency Locality Theory (DLT), or would serve as retrieval cues to help pre-build the RC structure, as predicted by the cue-based retrieval theory. Results from a visual world eye-tracking experiment and a self-paced reading showed that Chinese comprehenders are able to use BEI cues and the mismatching classifier (albeit to a less extent) to pre-build RC structure, providing support for the cue-based retrieval theory.

Keywords: relative clause; classifier; passive marker; sentence comprehension

Downloaded by [Peking University] at 07:40 19 October 2014

In recent years, there has been a growing interest in cross-linguistic research in the field of sentence processing, mainly because it can provide new perspectives to examine theories of language comprehension. One key question that the field endeavours to answer is how comprehenders of typologically different languages build up expectations about upcoming linguistic structures based on the input that has been seen or heard so far, and arrive at ultimately correct meaning representations by rapidly using various sources of information, including syntactic rules, thematic role assignment, real-world plausibility and discourse contexts (e.g. Altmann & Kamide, 1999; Altmann & Steedman, 1988; Gibson & Pearlmutter, 1998; Kamide, Altmann, & Haywood, 2003; Tanenhaus & Trueswell, 1995; Trueswell, Tanenhaus, & Garnsey, 1994). Recent research, in particular, has been interested in how comprehenders constrained by limited resources maintain in working memory linearly presented bottom-up word strings that are yet to be integrated into the target structure, while making top-down predictions proactively about possible structures before the sentence ends (e.g. Gibson, 1998, 2000; Levy, 2008; Konieczny 2000; Staub & Clifton, 2006; Van Dyke & Lewis, 2003; Vasishth & Lewis, 2006).

In this paper, we focus on two prominent working memory-based accounts that are currently under debate: the Dependency Locality Theory (DLT, Gibson, 1998, 2000) and the cue-based retrieval theory (Van Dyke & Lewis, 2003; Van Dyke & McElree, 2011; Vasishth & Lewis, 2006). Both theories contain some notion of expectation or anticipation during online sentence processing; both can account for parsing and reanalysis processes of ambiguous as well as complex (but potentially unambiguous) sentences (Gibson, 2000, pp. 115–119; Van Dyke & Lewis, 2003; but see Gennari & MacDonald, 2008 for discussions of structural ambiguity involved in English object-extracted relative clauses (RCs)); and both have supporting evidence from certain types of languages (i.e. head-initial and head-final). This paper aims to investigate how well these two theories extend to Mandarin Chinese, a language with mixed word order, and to distinguish their adequacy using Chinese prenominal RCs. Below we first review these two theories.

The DLT (Gibson, 1998, 2000) posits that human parsing systems consume computational resources available in working memory while keeping track of syntactic heads over a linear distance between a head and its dependents. Memory costs are measured by two metrics[1]: (1) the *storage cost* of maintaining syntactic heads required to complete a phrase-structural dependency (i.e. syntactic predictions), and (2) the *integration cost* of integrating a current word into an existing structure. These

*Corresponding author. Email: fywu@shisu.edu.cn (FW); xz104@pku.edu.cn (XZ)

© 2013 Taylor & Francis

Downloaded by [Peking University] at 07:40 19 October 2014

two metrics have been recast in various forms in the recent development of the DLT, for instance, the predicted syntactic head hypothesis is an equivalent counterpart for storage costs (Nakatani & Gibson, 2008, 2010), and so is the discourse-referent-based decay account for integration costs (Fedorenko, Piantadosi, & Gibson, 2012). But the general predictions that they make remain the same, i.e. the higher the number of predicted syntactic heads stored in working memory, the greater the computational difficulty; and the longer the linear distance between a head and its dependent in terms of intervening discourse referents (specifically, nouns and verbs), the greater the integration cost in establishing the dependency. Directly relevant to our paper, it is worth noting that the DLT, in particular its storage-cost metric, encodes the syntactic expectation in terms of 'storage' of the number of heads predicted.

In contrast, the cue-based retrieval theory (also called decay-based activation retrieval theory) (Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2011; Vasishth & Lewis, 2006) does not quantify the processing difficulty in terms of stored heads or some function of the linear distance between a head and its dependent. Rather, it describes real-time processing of an item in terms of its activation level: Speedup effects can be explained by increased activation resulting from repeated access and slowdown effects can be explained by the decay of activation over time. Retrieval interference is quantified by the extent to which retrieval cues provided by the phrasal head (e.g. the verb) can uniquely identify the correct dependent (e.g. its noun argument) from available candidates. A better fit between the intervening element and the head's retrieval cues will result in greater difficulty of retrieving the target element. But if intervening constituents can reactivate a decayed element, then integration of that element is facilitated. It is worth pointing out that this theory has a component of anticipatory parsing: if a target (lexical item or syntactic structure) is predicted by some events or retrieval cues, then accessing it will become easier than if it is not predicted.

Each of these two theories has supporting evidence from head-initial and head-final languages. The *integration cost* metric of the DLT has obtained substantial evidence from English (e.g. Gibson, 1998; Grodner & Gibson, 2005), showing that as the number of intervening discourse referents increases, more processing difficulty will ensue. Evidence for the integration cost metric has also been recently observed in Chinese prenominal RCs (Gibson & Wu, 2013). Support for the *storage cost* metric comes from Japanese nested structures (Nakatani & Gibson, 2008, 2010) and Chinese prenominal RCs (Hsiao & Gibson, 2003). The cue-based retrieval theory has been supported by English sentences involving filler-gap dependencies (Van Dyke & Lewis, 2003; Van Dyke & McElree, 2011; Hofmeister, 2011) or subject-verb dependencies (Bartek, Lewis, Vasishth, & Smith, 2011), Hindi nested structures (Vasishth, 2003; Vasishth & Lewis, 2006), and German nested structures (Konieczny, 2000).

One way to evaluate the DLT and the cue-based retrieval theory is to examine their predictions for a common construction in a language with different word order patterns than English. We chose to investigate RCs in Mandarin Chinese for the following two reasons. First, RCs are a useful phenomenon for cross-linguistic comparisons, and perhaps are the most studied (e.g. Keenan & Comrie, 1977; MacWhinney & Pléh, 1988). A large body of literature testing the predictions of the two theories reviewed above comes from RCs (e.g. Fedorenko, Piantadosi, & Gibson, 2012; Hsiao & Gibson, 2003; Nakatani & Gibson, 2010; Van Dyke & Lewis, 2003; Vasishth & Lewis, 2006). Second, Mandarin RC construction is particularly interesting given its typological uniqueness in combining (S)VO word order and (head)noun-final properties (Dryer, 1992; Greenberg, 1963), inherent ambiguity arising from mixed word order (e.g. Gibson & Wu, 2013; Lee, 2006; Lin, 2006; Wu, Kaiser, & Andersen, 2012), and lack of morphological case to clearly mark the clausal boundary. Consider the Chinese object-extracted RC (ORC) modifying the matrix object NP *chuanghu* 'window' in (1).

(1) Xiaomei yikai-le [$_{RC}$ zuqiu dasui __ de] chuanghu
Xiaomei **move-**ASP football break DE **window**
'Xiaomei **moved away the window** [$_{RC}$ that the football broke __]'.

In (1), the word order in the matrix clause and in the RC is SVO as in English, but unlike in English, the RC precedes the head noun *chuanghu* 'window'. As is widely known in RC processing, a comprehender needs to recognise the RC boundary and establish a dependency – commonly known as a filler-gap dependency (e.g. Clifton & Frazier, 1989; Frazier & d'Arcais, 1989) – between the head noun (i.e. the filler, 'window') and the empty nominal within the RC (i.e. the gap). However, the adnominal marker, DE, as an indispensable particle in RC (roughly equivalent to the relative pronoun 'that' in English), does not occur until right before the sentence-final head noun. In addition, the distance between the main verb *move* and its direct object 'window' (i.e. the head of RC) is greater in Chinese than its English counterpart, wherein the main verb is immediately adjacent to its argument. All this may render Chinese head-final RC parsing to appear particularly difficult or 'inefficient' (Hawkins, 2004, p. 145). In fact, given that a comprehender prefers to choose a structurally simple analysis among a number of possible analyses (e.g.

Downloaded by [Peking University] at 07:40 19 October 2014

Frazier, 1987; Kimball, 1973), the RC-internal subject (*zuqiu* 'football') is very likely to be initially misanalysed as the direct object of the main verb (*yikai* 'move') when the sentence with an object-modifying ORC (1) is presented in isolation. Thus, the comprehender might experience a garden-path effect upon encountering the next RC-internal verb (*dasui* 'break'), which is indicative of an ORC structure as the most probable continuation.

One possible way to modulate the garden-path effects in sentences like (1) is to provide cues prior to the head noun such that they could potentially increase the expectations for an RC, hence facilitating the retrieval of the target noun 'window'. For instance, if a determiner phrase consisting of a demonstrative (Dem) and a classifier (CL) *shan*, which can occur together with *chuanghu* 'window', is added to the left edge of the ORC as in (2), the comprehender might be less likely to garden-path upon encountering the embedded noun *zuqiu* 'football'. The logic is that the classifier *shan* can only modify a flat entity that can be open and closed, specifically a window or door, whereas the immediately adjacent noun ('football') is round and cannot be open or closed flexibly. Thus, the Chinese comprehender might use the demonstrative-classifier sequence to predict that a noun is upcoming that has to match the classifier *shan*. A noun which mismatches the immediately preceding classifier cannot be integrated as the object of the main verb 'move', and would thus be interpreted as the beginning of an ORC, which can be mentally constructed prior to the adnominal marker DE.

(2) Xiaomei yikai-le na-**shan** [$_{RC}$ zuqiu dasui __ de] **chuanghu**

Xiaomei move-ASP that-CLwindow football break DE window

'Xiaomei moved away the window that the football broke __'.

Building upon this line of reasoning, if an additional passive (clause) marker BEI is added immediately following the demonstrative-classifier sequence as in (3), it is very likely for the comprehender to posit a clausal boundary upon encountering BEI.

(3) Xiaomei yikai-le na-**shan** [$_{RC}$__ **bei** zuqiu dasui __ de] **chuanghu**

Xiaomei move-ASP that-CL$_{book}$ PASS football break DE window

'Xiaomei moved away the window that was broken by the ball'.

The logic is that the presence of BEI minimally requires a patient (in this case, *chuanghu* 'window'), a verb (*dasui* 'break') and an optional agent or (in this case) instrument (*zuqiu* 'football'), in order to complete the semantic closure of a passive event. Due to Chinese grammar, the noun following BEI, if present as in (3), has to be interpreted as the agent/instrument of an action which, in this case, is different from the main verb *yikai* 'move'. Thus, the Chinese comprehender may use this passive marker BEI, together with the preceding classifier mismatch *shan* that implies a semantically congruent head noun yet to come, to construct an RC[2] structure, which is the only possible continuation to end the sentence.

Whereas the presence of mismatching classifier and BEI cues in (2–3) are likely to help comprehenders to posit a clausal boundary early and to anticipate an RC, those pre-RC heads also entail additional structural complexity compared to (1). At least three kinds of dependencies are to be established in order to arrive at a meaning representation of sentences such as (2): the filler-gap dependency, the classifier-(head) noun dependency and the long-distance dependency between the main verb 'move' and its argument (i.e. the head noun) 'window'. All these dependencies are to be completed at the head noun in the sentence-final position. Questions then arise: How would the processes of constructing the RC structure or retrieving the head noun be affected by these pre-RC heads? Would these processes be facilitated because the RC structure is successfully cued, or otherwise be burdened because more dependencies are yet to be completed? Specifically, could the lexical cue of mismatching classifier alone, or in conjunction with the syntactic cue of BEI, help Chinese comprehenders to speed up the comprehension of the RC structure?

Existing work on Chinese has yet to give clear answers to those questions above. Consider the mismatching classifier cue first. While evidence from Japanese demonstrates that consistent head-final properties and overt case marking allow comprehenders to use classifier mismatch to project a head noun rather than attaching the classifier to a local noun (Phillips & Lau, 2004; Yoshida, 2006; Yoshida, Aoshima & Phillip. 2004), research on Chinese has not robustly shown that Chinese comprehenders are able to solidly construct the RC structure upon encountering the classifier-noun incongruity. In fact, existing evidence suggests that mismatching classifiers may not effectively build the RC parse in the absence of discourse contexts, due to (1) the infrequency of mismatching classifier-noun sequences (Wu, 2011), (2) comprehenders' propensity to treat such mismatching sequences in written texts as 'typos' (Hsu, 2006) and (3) the structural complexity of RCs spanning a distance between a classifier mismatch and its hosting noun (Hsu, Phillips, & Yoshida, 2005). We review each of them below.

Regarding the frequency of occurrence, a recent examination of the Chinese Treebank Corpus 5.0 showed that the classifier-noun mismatch configuration rarely occurred in written news texts (Wu, 2011). Out of the 10 tokens of ORCs with pre-RC classifiers, 3 contained classifier-noun match configuration (e.g. *nei-zhi zuqiu dasui de chuanghu*… '[$_{RC}$ that-CL$_{ball}$ football broke t_i DE] window$_i$', 'the window that the football broke'), 5 had lexically and phonetically empty RC-subjects (e.g.

Downloaded by [Peking University] at 07:40 19 October 2014

nei-shan dasuide chuanghu '[$_{RC}$ that-CL$_{window}$ (e$_i$) broke t_j DE] window$_j$', 'the window that someone broke'), and 2 contained some additional lexical item (similar to a passive marker) intervening between an otherwise 'mismatching' classifier and the RC-subject. Such distributional patterns suggest that Chinese speakers or writers are biased to put a matching classifier next to its hosting noun, and to avoid cases in which a classifier is adjacent to a semantically incongruent (i.e. mismatched) noun.

The corpus results may partially explain the response patterns of an offline test conducted by Hsu (2006). When participants were asked to give written completions for sentence fragments with classifier-noun mismatches (e.g. *na-ben zuojia*… 'that-CL$_{book}$ writer…'), 50% (108/216) of the completions were ungrammatical simple clauses, 44% (94/216) were RCs and 6% were NP-deletion responses. Overall, the chances were roughly equal between producing a structurally simpler monoclausal structure at the expense of a well-formed grammaticality and producing an RC that is ultimately grammatical but structurally more complex. This suggests that at least to some Chinese speakers, the mismatch between a classifier and its following noun is more likely to be considered as an error, thereby satisfying the constraint of local classifier-noun congruity.

While the production-based research reviewed above may not speak directly to the online comprehension processes, existing work on the role of the mismatching classifier cues in Chinese ORC processing has also yielded mixed results (Hsu et al., 2005; Hsu, Hurewitz, & Phillips, 2006; Wu, Haskell, & Andersen, 2006). In a series of self-paced reading studies, Hsu et al. (2005) found that the mismatching classifier induced a long-lasting lexical disruption effect at the embedded noun region, with no facilitatory effects at the head noun. It is worth noting that their stimuli always contained adverbial phrases that intervened between the embedded RC-subject and the head noun, as in 'that-CL$_{vehicle/*human}$ college student [$_{ADV}$ often with careful attitude] maintain DE motorcycle…', 'that motorcycle that the college student maintained often and with a careful attitude'. By the DLT, (1) the additional phrasal heads postulated due to the intervening adjuncts and (2) the new discourse referent (i.e. the NP within the prepositional phrase) might have rendered a high *storage* and *integration* cost. By the cue-based retrieval theory, the activation of the head noun might decay ever since it was first cued by the mismatching classifier; because it is not reactivated over time and its retrieval at the RC-verb 'maintain' possibly fails. Thus, both theories can explain Hsu et al.'s findings for no effects of mismatching classifiers on the head nouns when ORCs are presented in isolation.

In an eye-tracking study, Wu et al. (2006) questioned the plausibility of expecting participants to utilise the classifier mismatch cue when target sentences were presented in isolation, given that an RC is only necessary in a context where comprehenders need to single out one unique referent out of a set. Building upon Wu et al. (2006) and following the referential theory of Altmann and Steedman (1988), Hsu et al. (2006) created discourse contexts in a self-paced reading task. They found that at the head noun region, mismatching classifiers greatly facilitated reading times (RTs) when preceded by a 2-referent context (with two identical referents, e.g. *two motorcycles*), but not when preceded by a 1-referent context (with two different referents which do not share the same classifier, e.g. *a motorcycle* vs. *a computer*). This finding suggests that given RC-biasing contexts, Chinese comprehenders may use the mismatching classifier to recognise the left edge of the RC.

Turning now to the role of BEI, we are only aware of one published self-paced reading study that directly tested the role of BEI as well as mismatching classifiers in processing Chinese subject-modifying RCs. Wu, Kaiser, and Andersen (2009) compared the effects of these cues against the cases where mismatching classifiers, or BEI, or no cues (i.e. bare nouns) were provided in RCs presented in null contexts. The results showed facilitative effects of BEI and mismatching classifiers compared with no-cue conditions within the RCs.

In summary, existing work on the role of mismatching classifiers in ORCs has shown little facilitation at the head noun when sentences are presented in isolation. But such facilitations have been found in the presence of 2-referent contexts, possibly because RCs usually serve to select a particular referent from a set. It is worth noting that most existing studies not only have their stimuli presented in null contexts, but also have focused on RCs that modify sentential subjects (i.e. subject-modifying RCs). Thus, in structurally complex object-modifying (O)RCs, which may additionally induce garden-pathing at the embedded noun, it remains unclear what the processing consequences are for mismatching classifiers – both alone and in conjunction with BEI – against the predictions of the two theories.

To further investigate these issues, we conducted two experiments with essentially similar stimuli, one using visual world eye-tracking and the other self-paced reading. We focused solely on Chinese object-modifying RCs with a schematic structure of '(S) V [$_{RC}$ …] headN'. All RC sentences were preceded by discourse contexts, but unlike Hsu et al. (2006), we only used 2-referent contexts (i.e. two identical or similar-kind referents) to satisfy pragmatic presuppositions behind the use of RC.

Experiment 1: visual world eye-tracking

Using the visual world eye-tracking paradigm with spoken sentences, Experiment 1 aims to investigate the effects of pre-head cues that vary in predictive value during

Figure 1. Sample visual display of classifier-mismatch conditions in Experiment 1.

(4) **Match/no-BEI Condition:**
'There are two vases. A football breaks this one. A brick breaks this one'.
Critical sentence:

Dianji	[na-**zhi**	**zuqiu**	dasui	t	de] **huaping**$_i$
click-on	that $CL_{football/vase}$	football	break	t_i	DE vase$_i$

'Click on the vase that the football broke'.

(5) **Match/BEI Condition**

Dianji	na-**zhi**	[t_i bei	**zuqiu**	dasui	de] **huaping**$_i$
click-on that	$CL_{football/vase}$	t_i PASS	football	break	DE vase$_i$

'Click on the vase that was broken by the football'.

(6) **Mismatch/no-BEI Condition**:
'There are two windows. A football breaks this one. A brick breaks this one'.
Target trial:

Dianji	na-**shan**	[zuqiu	dasui	t_i	de] **chuanghu**$_i$.
click-on	that CL_{window}	football	break	t_i	DE window$_i$

'Click on the window that the football broke'.

(7) **Mismatch/BEI Condition**

Dianji	na-**shan**	[t_i bei	zuqiu	dasui	de] **chuanghu**$_i$.
click-on	that CL_{window}	t_i PASS	football	break	DE window$_i$

'Click on the window that was broken by the football'.

incremental parsing of Chinese head-final RC, in particular the time course of effects associated with syntactic expectations, lexical disruptions and retrieval interference. Visual world eye-tracking is known to provide time-locked information regarding what induces parsing difficulty (e.g. Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Tanenhaus & Trueswell, 2006). One additional benefit of visual world eye-tracking is its aural presentation mode, which may render the classifier-noun incongruity apparent to native speakers' ears, making it unlikely for participants to completely ignore the incongruence or treat it as an 'error' as reported in earlier work using written presentation mode (Hsu, 2006).

In this design, participants viewed pictures and listened to four sentence trials while their eye movements were monitored. On each trial, the first three (leading-in) sentences created a 2-referent context, followed by the critical instruction to click on a target picture (see below). The visual display was composed of four pictures (Figure 1), two of which correspond to the head NP ('window') in the critical instruction, one corresponds to the embedded noun ('football') and one corresponds to the alternative noun ('brick') in one of the leading-in sentences. Given that both the embedded noun and the alternative noun were linked to the head NP through the verb in the leading sentences, participants have to rely on the information provided by an RC to identify which of the two duplicated pictures was being referred to in the instruction. Filler trials were added to counterbalance the overall chance for the sentence to end as an RC or as a main clause, that is, a direct-object parse (e.g. 'Click on the brick). Thus, both parses were likely to be activated prior to the presentation of the critical instruction.

We crossed Classifier (match vs. mismatch) and BEI (present vs. absent), yielding four conditions as in (4–7): Match/no-BEI, Match/BEI, Mismatch/no-BEI and Mismatch/BEI. Note that in the Match conditions (4–5), the classifier *zhi* matched the embedded noun *zuqiu* 'football' as well as the head noun of the RC *huaping* 'vase'; in the Mismatch conditions (6–7), the classifier *shan* mismatched the embedded noun *zuqiu* 'football' but matched the head noun of the RC (e.g. *chuanghu* 'window').

Predictions made by the DLT

Table 1 illustrates the specific prediction of the DLT's storage-cost metric measured in memory units (MUs) for the four sentences at each word, with the crucial RC regions shaded in grey. As shown in the last row of each experimental condition, the storage-cost metric predicts less storage costs at the embedded noun position in the Match/no-BEI condition than the other three conditions, and crucially, no processing differences are predicted to arise in all other regions within the RCs.

To see how the MUs are computed, we use the Match/no-BEI condition for illustration. The DLT assigns an MU for each syntactic head predicted that is needed to complete dependencies to be built. Upon encountering the verb ('click-on'), a Chinese comprehender needs to predict a bare noun (N) minimally in order to build a main clause. At the next available demonstrative-classifier sequence ('that-zhi'), the classifier 'zhi' leads the comprehender to predict a semantically congruent noun. Because the next available noun ('football') matches the preceding classifier 'zhi', comprehenders can establish the classifier-noun dependency and the verb-argument dependency by directly attaching the classifier into the local noun. Thus, at this word, no heads need to be predicted, yielding zero MUs. This monoclausal direct-object structure has to be disambiguated into an ORC at

Downloaded by [Peking University] at 07:40 19 October 2014

Table 1. Word by word predictions of the storage-based metric in DLT.

Match, no-BEI (ORC)					Input word			
		Dianji	[na-zhi		zuqiu	dasui	de]	huaping
		Click-on	that-CL		football	broke	DE	vase
Storage cost	Syntactic heads predicted	N	N			DE, headN	headN	headN
	MU	1	1		0	2	1	0
Match, BEI (SRC)					Input word			
		Dianji	na-zhi	[bei	zuqiu	dasui	de]	huaping
		Click-on	that-CL	PASS	football	broke	DE	vase
Storage cost	Syntactic heads predicted	N	N	N,V,DE,headN	V,DE,headN	DE, headN	headN	
	MU	1	1	4	3	2	1	0
Mismatch, no-BEI (ORC)					Input word			
		Dianji	na-shan		[zuqiu	dasui	de]	chuanghu
		Click-on	that-CL		football	broke	DE	window
Storage cost	Syntactic heads predicted	N	N		V,de,headN	DE, headN	headN	
	MU	1	1		3	2	1	0
Mismatch, no-BEI (SRC)					Input word			
		Dianji	na-zhi	[bei	zuqiu	dasui	de]	chuanghu
		Click-on	that-CL	PASS	football	broke	DE	window
Storage cost	Syntactic heads predicted	N	N	N,V,DE,headN	V,DE,headN	DE, headN	headN	
	MU	1	1	4	3	2	1	0

Downloaded by [Peking University] at 07:40 19 October 2014

the next available verb ('broke'), thus two heads (DE, head noun) are predicted.

In terms of integration cost, the DLT specifies that integration costs are incremented by the number of new contentful discourse referents (i.e. nouns and verbs) that intervene between the head and dependent. In our study, the two intervening referents (i.e. the noun 'football' and the verb 'broke') are present in the context, thus no extra work is involved in constructing discourse-old referents (see Gibson's stance on this issue in Fedorenko et al., 2012, p. 476). This means little or no integration costs at the head noun in all the three kinds of dependencies.

Taken together, the predictions made by the DLT mainly come from the storage-cost metric. Related to our eye-tracking experiment, the DLT predicts that more predicted heads in the three conditions other than the Match/no-BEI condition will have destructive effects on anticipatory looks to the target picture (i.e. the head NP) when the embedded noun is heard. Thus, participants should delay in programming looks to the target picture until probably after they hear the verb-DE. Because this is also the disambiguation point for the Match/no-BEI condition, no differences in the proportion of looks to the target picture are predicted across the four conditions from the verb-DE[3] up until the end of the sentence. We will return to this point when we discuss later the saccades of moving out of embedded NP and moving into target pictures.

Predictions made by the cue-based retrieval theory

We now turn to the predictions of the cue-based retrieval theory. Following Van Dyke and Lewis (2003) and Vasishth and Lewis (2006), we assume a left-corner parser that builds up syntactic trees for sentences (4–7) as shown in Figures 1–3. To construct the correct RC parse, the parser needs to successfully retrieve the head NP node ('vase' in Match conditions and 'window' in Mismatch conditions). According to this theory, a target item's activation level is determined by (1) its usage history (i.e. the number of times it has been retrieved), (2) time-based decay since its retrieval and (3) interference from intervening items that match the to-be-retrieved target item (i.e. similarity-induced interference).

In terms of retrieval history of the head NP, across the four conditions the head NP is first activated proactively (as a potential argument) at the matrix verb, and then repeatedly reaccessed at the classifier and embedded verb.

Downloaded by [Peking University] at 07:40 19 October 2014

It is also reaccessed at BEI in the BEI conditions. Thus, the head NP has three boosts in activation in the no-BEI conditions, and four boosts in activation in BEI conditions. Note that at the retrieval cue of the classifier, the matching classifier 'zhi' is associated with three semantically congruent NPs (i.e. two vases and a football), whereas the mismatching classifier 'shan' is associated with only two NPs (i.e. two windows). Thus, the strength of association from the classifier cue to the head NP is reduced in the Match conditions (due to more competitors) than the Mismatch conditions.

In terms of interference from the intervening embedded NP ('football') that dampens activation of the head NP, let's examine each of three types of dependency in turn. Regarding the gap-filler dependency, the embedded NP is intervening in the BEI conditions, but not in the no-BEI conditions. Given that the passive marker BEI as a syntactic cue can straightforwardly slot the embedded NP into the non-head agent/instrument node and the to-be-retrieved head NP into the RC-subject node (see Figure 2), interference from the embedded NP is likely to be minimal in the two BEI conditions. Thus, no processing differences are expected in processing gap-filler dependency.

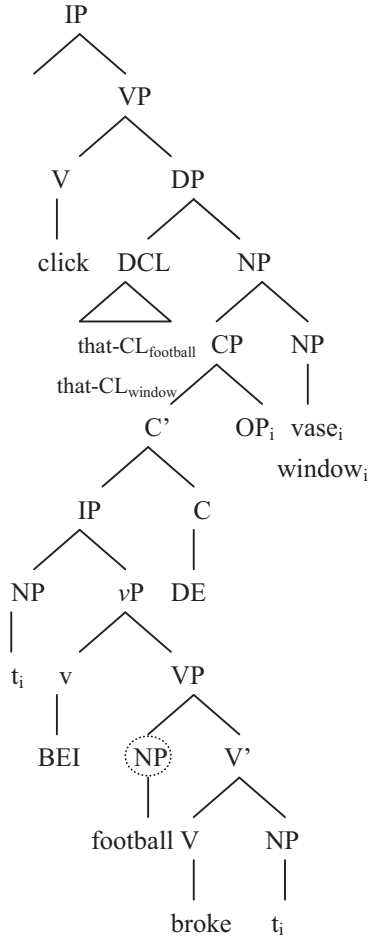

Figure 2. RC parse with the intervening embedded NP in the BEI conditions.

Regarding the classifier-noun dependency, the embedded NP in the Match/no-BEI condition is already attached into the preceding classifier, forming a DP node (see Figure 3b), thus it does not interfere in the retrieval of the head NP. There is little or no interference from the intervening embedded NP in the two BEI conditions, either, for the same reason as stated above in the case of gap-filler dependency. But in the Mismatch/no-BEI condition, the embedded NP (see Figure 4b) is likely to interfere in the retrieval of the head NP because of shared features such as [-animate].

Regarding the verb-argument dependency, across all four conditions the embedded NP is intervening between the matrix verb and the to-be-retrieved head NP. But its interference is likely to be minimal in the two BEI conditions because the presence of the syntactic marker BEI (1) should block any possible attachment of the embedded noun 'football' into the preceding classifier and (2) will increase the retrieval of the head NP as the subject of the passive event (see Figure 2). Thus, the subsequent retrieval of the head NP at the RC-verb ('broke') should be faster in the BEI conditions than the no-BEI conditions.

In the Match/no-BEI condition, the interference from the embedded NP inhibiting the retrieval of the head NP is likely to be all-or-none because syntactic reanalysis involved might not be difficult to recover. Specifically, the parser may initially attach the embedded noun into the preceding classifier *zhi*, forming a DP, and then attach this DP into the existing parse tree, thereby completing the verb-argument dependency (Figure 3a). This direct-object parse may completely inhibit the retrieval of the head NP because such retrieval is rendered unnecessary. However, upon encountering the next available disambiguating verb 'broke', the parser must break the link between the main verb 'moved away' and the DP node 'that-CL football' in order to attach 'broke' into the existing parse. The DP node 'that-CL football' (which was the object of the main verb 'moved-away') must now be *reattached* as the subject of the RC headed by a noun yet to be retrieved (Figure 3b). Reanalysis normally would incur processing difficulty, but given that the disambiguating verb is immediately adjacent to the ambiguous NP, recovery from garden-pathing may be successful at the RC-verb 'broke' (see Van Dyke & Lewis, 2003, p. 291 and relevant work wherein). But it is probable that 'reattachment' (i.e. repositioning of the DP node downstairs in the tree) could be difficult.

In the Mismatch/no-BEI condition, the intervening embedded NP might cause interference in retrieving the head NP at the RC-verb. The parser may use the semantic clash induced by the classifier-noun incongruity to *detach* the embedded NP from the right branch of the DP node (Figure 4a). But to *reposition* this 'stranded' NP

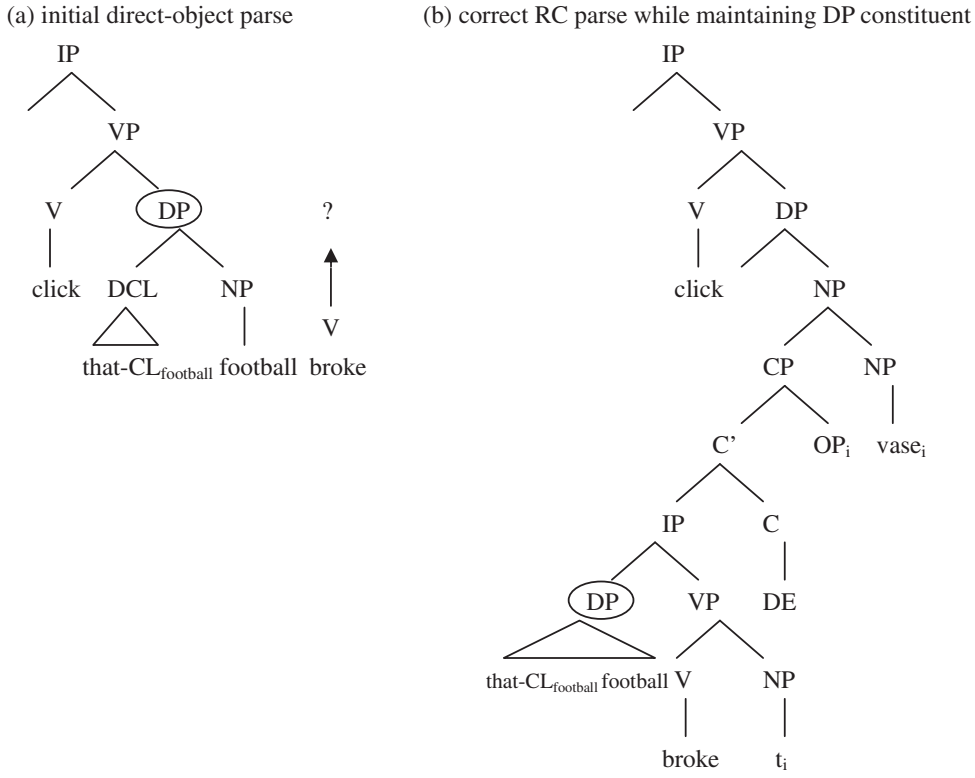

Figure 3. (a) Initial analysis of the Match/no-BEI sentence as the incorrect direct-object parse. (b) Final RC parse.

Downloaded by [Peking University] at 07:40 19 October 2014

appropriately into the existing parse (Figure 4b) might be costly because of similarity-induced inhibitions from the recently retrieved/activated target NP by the retrieval cue of mismatching classifier. At the retrieval cue of RC-verb 'broke', the quality of representations of the retrieved head NP is in turn likely to be modulated by the interference from the preceding embedded NP.

Thus, in processing verb-argument and/or classifier-noun dependencies, the embedded NP is likely to induce varying interference in the two no-BEI conditions, but such interference is likely to be minimal in the two BEI conditions.

Taken together, the cue-based retrieval theory predicts a main effect of BEI and an interaction at the embedded NP and RC-verb. Specifically in the Mismatch/no-BEI and two BEI conditions, there should be early looks to the target picture prior to the onset of the disambiguating verb, because the mismatching classifier and BEI as cues should deactivate the direct-object parse and pre-activate the head noun in the RC parse, and such expectations for the upcoming head noun should be reinforced as retrieval cues unfold in time due to repeated access.

Method

Participants

Twenty-eight native speakers of Chinese from Peking University participated in the experiment in exchange for 20 RMB. Their mean age was 21. They all had normal or corrected-to-normal vision.

Materials

Twenty-four experimental items were constructed, each consisting of a set of four sentences, along with a set of four pictures. In the critical sentence, the semantic congruity of classifiers (match or mismatch) and the passive marker BEI (presence or absence) were crossed, resulting in four conditions, as shown in (4–7). The head nouns in the four conditions were matched in terms of lexical (log) frequency and word length.

Auditory stimuli. The sound files that participants heard while viewing the visual displays were produced by a female native speaker of Mandarin naive to the purpose of the experiment, and were digitally recorded with a PC using Adobe Audition (16 bits; 44,100 Hz) in a sound-proof chamber.

Given that the classifier and the embedded noun may form a DP constituent in the Match/no-BEI condition but belong to different clauses in the Mismatch/no-BEI condition, it is plausible that these two elements might be realised differently in terms of prosody. Existing phonetic work in Mandarin Chinese as well as in other languages indicates that prior to prosodic phrasal boundaries, the preceding syllable is usually lengthened, and a pause inserted (e.g. De Pijper & Sanderman, 1994; Li & Yang, 2009; Wang, Yang, & Lue, 2004; Wightman, 1992).

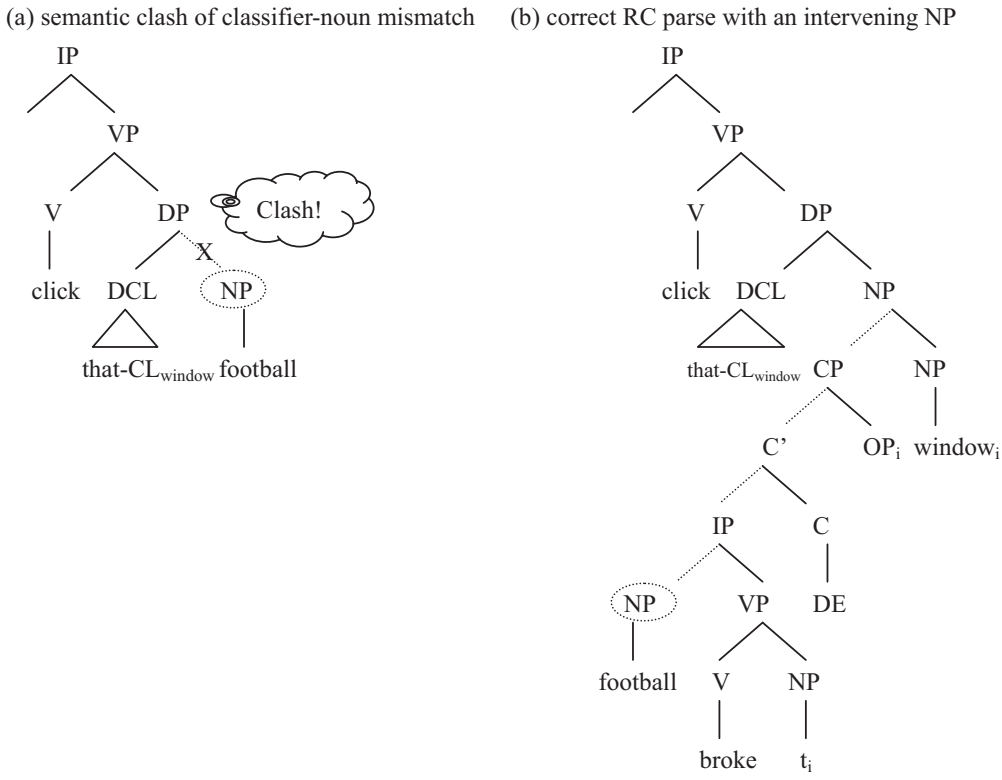

Figure 4. (a) Initial semantic clash of the classifier-noun mismatch in the Mismatch/no-BEI condition. (b) Final RC parse.

Consistent with the prosodic distinctions reported in the literature, in our stimuli, the average duration of mismatching classifiers (336 ms) was numerically longer than that of matching classifier (320 ms), with a numerically longer pause inserted before the embedded noun in the Mismatch/no-BEI condition (7 ms) than in the Match/no-BEI condition (4 ms). In addition, the average duration of the embedded noun was numerically shorter in the Mismatch/no-BEI condition (584 ms) than in the Match/no-BEI condition (596 ms), perhaps to compensate for the preceding syllabic lengthening effect. But none of these differences reached significance: for classifiers: $t(46) = 1.2$, $p = 0.24$; for pauses: $t(46) = 0.7$, $p = 0.49$; for embedded nouns: $t(46) = -0.39$, $p = 0.7$. Thus, other than the factors manipulated in our experiment, it is unlikely that the acoustic stimuli contained prosodic cues that were distinctly salient enough for participants to use during online processing.

Visual stimuli. The pictures which accompanied each set of sentences were obtained from the Internet. All pictures were scaled to the same size. If the picture was larger in one dimension than the other, white space was added on the sides or top and bottom to make it approximately square. The position of the target picture (i.e. the one that the participant was instructed to click on) was counterbalanced across items, so that it occurred equally often in each corner of the screen. However, the two duplicated pictures (e.g. 'windows' in the Mismatch conditions or 'vases' in the Match conditions) always occurred next to each other, either in the top row or the bottom row. The position of the non-duplicated pictures (e.g. the embedded noun 'football' or the distractor 'brick') was also counterbalanced, so that half the time the picture on the left was mentioned first, and half the time the picture on the right was mentioned first. The position of the agent/instrument picture (e.g. 'football' or 'brick') was always vertical, but never diagonal, to the position of the patient picture (i.e. one of the duplicated pictures). A sample visual display is given in Figure 4.

There were 24 sets of target images, each in two versions. One version corresponded to the two Match conditions and the other to the two Mismatch conditions.

Fillers. Thirty-six filler items were constructed in addition to the experimental items. Twenty of the fillers had the same structure as the experimental items, except that the critical sentence contained a simple noun phrase, rather than one with an RC, as in (8). This simple NP always referred to one of the agents/instruments of the action (e.g. the equivalent of the brick or ball).

(8) Sample Filler Item: direct object noun phrase

'There were two watermelons. The zebra ate this one. The rhinoceros ate this one.

Click on **the zebra**'.

Downloaded by [Peking University] at 07:40 19 October 2014

The other 16 fillers were also similar in structure to the experimental items, except that in the context sentences, the duplicated picture depicted the agent or instrument rather than the patient of the action. In the fourth sentence, eight fillers contained a subject-extracted RC rather than an object-extracted RC, as in (9a), and eight fillers contained a simple noun phrase as in (9b).

(9) Sample Filler Item:

a. Subject-extracted RC
'There are two airplanes. This one flies over a snow mountain. This one flies over a desert. Click on **the airplane that flew over the snowy mountain**'.

b. Direct object noun phrase
'There are two fish rods. This one has a caterpillar on it. This one has an earthworm on it. Click on **the earthworm**.

Thus, of all the 60 sentences a participant heard, 32 contained an RC (12 ORCs; 20 SRCs including 12 passive relatives) and 28 contained a simple noun phrase.

Four stimulus lists were created using a Latin Square design. Each list contained one version of each experimental item, along with all 36 filler items, for a total of 60 items. The order of items was pseudo-randomised for each list, such that all target items were separated by at least one filler, and no target items occurred in the first two trials for each list.

Procedure

Participants were seated with their heads positioned on a chin rest 80 cm from the screen, and held a game pad in their hands. Eye gaze was recorded with an Eyelink 2K eye-tracker (SR Research, sampling rate 2000 Hz). Participants listened to the pre-recorded stimuli while viewing colour pictures displayed on the computer screen. They were instructed to use one of the four buttons on the game pad to click on the region of the picture according to the description of the instruction.

The visual and auditory stimuli were presented on a PC with a 21-inch CRT (cathode ray tube) monitor, using ExperimentBuilder software (SR Research). Auditory stimuli were played over SONY MDR-V900Hd headphones. Calibration was done using a 9-point calibration display. Participants completed three practice items before the experiments began.

Each trial began with the presentation of the four stimulus pictures on a grey background. When the second sentence was played, a yellow outline appeared surrounding the instance of the duplicated object that was being referred to (in Figure 1, this would be 'the window' on the left), and then was removed. When the third sentence was played, a yellow outline appeared surrounding the other instance of the duplicated object ('the window' on the right), and then was removed. Prior to the presentation of the critical sentence, a high-frequency warning tone was presented, together with a fixation cross at the centre of the screen for 1000 ms. Participants were required to fixate at the cross until the cross disappeared, at which time the critical sentence was presented. The display remained on the screen until the participant pressed on the game pad the button corresponding to one of the pictures on the visual display, at which point the next trial began. Drift corrections were performed every trial. All calibrations and recordings were based on the position of the left eye.

Coding

Fixations were classified as falling into one of four regions. The *target* region consisted of the picture that the participant was instructed to click on. Note that this picture was always duplicated; the other instance of the picture constituted the *contrast* region. The *embedded* region consisted of the picture corresponding to the embedded noun. The fourth picture, corresponding to the alternative noun in the leading-in sentence, constituted the *distractor* region.

Results

Response accuracy

Averaging across all target and filler trials, all participants clicked on 97.6% of the target pictures correctly on the game pad. On critical trials, the overall accuracy rate across participants was 97.9%. Experimental items with incorrect responses were eliminated from further analyses.

Eye movements

Due to the additional presence of BEI in two BEI conditions, the RC sentences are identical only after BEI. We therefore aligned our eye movement data by the onset of each individual embedded noun across conditions and marked it as zero. Given that it takes around 150–200 ms to programme an eye movement (Hallett, 1986; Matin, Shao, & Boff, 1993), we conducted statistical analyses on a series of 200 ms time slices, beginning at 200 ms after the onset of the embedded noun across the four conditions, and extending 1600 ms after the onset of the embedded noun (i.e. after the average onset of head nouns, which is 1269 ms). The same patterns of results were obtained when narrower time windows (e.g. 100 ms per slice) were used for statistical analysis

We used linear mixed-effects regression model with the lme4 package for the statistical language R (R Core Development Team, 2008) because in this model, participants and items are simultaneously entered as crossed random factors, and computation is relatively immune to missing data (Baayen, 2008; Baayen, Davidson, & Bates,

Downloaded by [Peking University] at 07:40 19 October 2014

Downloaded by [Peking University] at 07:40 19 October 2014

2008; Kliegl, Risse, & Laubrock, 2007). To evaluate the predictions made by the two theories, we focus on (1) the proportion of fixations to the embedded noun picture, to see whether direct-object analyses were entertained by participants, and (2) the proportion of fixations to the target-noun picture, to see whether RCs could be built early and how well RC representations were built (i.e. whether the intervening embedded NP interfered with the retrieval of the target NP). Viewing time (i.e. the duration of time for a participant to fixate on a particular picture) was measured in every 200 ms time slide and transformed in percentage (i.e. proportion of fixations = viewing time / 200).

In addition to the proportion of fixations, we also measure the probability of launching a saccade from or towards a picture in every 400 ms time window after the onset of the embedded noun. Two measurements were included in the statistical analyses. One is the probability of saccades towards the target picture regardless of where the saccades were from, and the other is the probability of saccades specifically from the embedded noun picture to the target picture. While the former can be viewed as reflecting whether anticipations of the target would be influenced by the auditory input, the latter is to answer the question of whether the syntactic or lexical cue discourages the direct object parse and hence strengthens the RC prediction. Both measurements were taken as binary variables in that they could represent either 'there was such a saccade' or 'there was no such a saccade (covering both saccades towards other areas and no saccade launched)' for each trial in a single time window. Estimates are from a generalised linear mixed model for these two variables, with crossed random effects for participants and items.[4]

Proportion of fixations to the embedded noun

Figure 5 shows the proportion of fixations to the embedded noun from the onset of the embedded noun across the four conditions. Given that the disambiguating point is around 200 ms after the onset of the verb, by then it should be clear that the direct-object analysis is incorrect, we only focused on the statistical analyses up to the verb (average onset of verb-DE = 585 ms; mean duration of verb-DE = 666 ms), i.e. 1000–1200 ms after the onset of the embedded noun. Table 2 presents the statistical results.

During the time slice from 200 to 400 ms after the onset of the embedded noun, there was a significant effect of classifier incongruity and an interaction, but no significant effects of BEI. Follow-up tests showed that mismatching classifiers induced a lower proportion of looks to the embedded noun than matching classifiers in both BEI conditions (11.6%; $b = 0.115$, SE = 0.019, $t = 6.2$, $p < 0.001$) and no-BEI conditions (5.6%; $b = 0.045$, SE = 0.018, $t = 2.5$, $p = 0.01$), and the magnitude of classifier incongruity effect was larger when BEI was present than when BEI was absent.

During the time slice from 400 to 600 ms post-embedded noun (i.e. prior to the onset of the verb, which is 585 ms after the onset of the embedded noun), there was a main effect of classifier incongruity. Mismatching classifiers induced a lower proportion of looks to the embedded noun picture than matching classifiers. There was also a main effect of BEI. The presence of BEI decreased proportion of looks to the embedded noun picture as compared with the absence of BEI. There was no interaction.

Similar results were also found from 600 to 800 ms post-embedded noun (i.e. 200 ms after the onset of the verb). There were main effects of classifier incongruity and BEI, but no interaction.

During the time slice from 800 to 1000 ms post-embedded noun when the verb was presented, there were no effects of classifier incongruity or BEI, but a marginal interaction. Follow-up tests revealed that a mismatching classifier led participants to be less likely to fixate on the embedded noun picture than a matching classifier only in

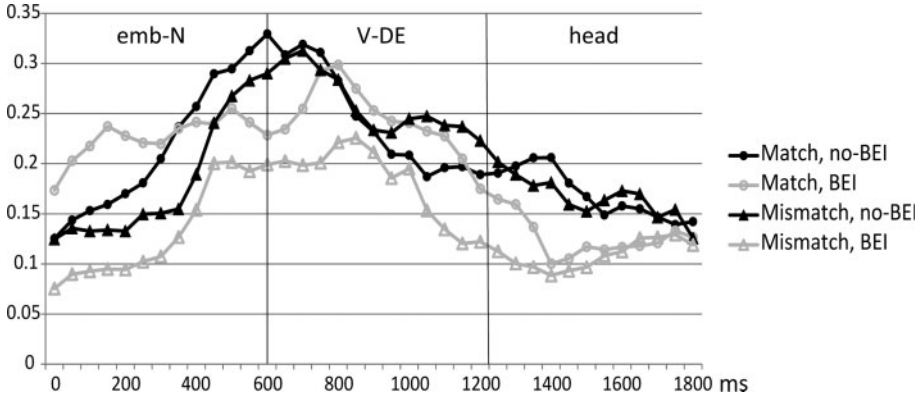

Figure 5. Proportion of fixations to the embedded noun picture in the four conditions in Experiment 1 (0 ms marks the average onset of embedded nouns across the four conditions).

Downloaded by [Peking University] at 07:40 19 October 2014

Table 2. Linear mixed-effects regression model for the proportion of looks to the embedded noun pictures after the onsets of embedded noun across the four conditions (Experiment1).

					CL incongruity					BEI					Interaction			
Time windows	M (%)	M, BEI (%)	MM (%)	MM, BEI (%)	Mean difference (%)	β	SE	t	p	Mean difference (%)	β	SE	t	p	β	SE	t	p
200–400 ms	20.2	22.2	14.6	10.6	8.6	0.079	0.013	5.9*	< 0.001	1.1	0.005	0.013	0.4	> 0.1	0.055	0.027	2.1*	0.04
400–600 ms	29.2	24.2	24.4	18.4	5.3	0.045	0.015	3.0*	0.003	5.5	0.052	0.015	3.5*	0.006	0.004	0.03	0.1	> 0.1
600–800 ms	31.8	25.1	29.9	19.8	3.6	0.033	0.016	2.1*	0.03	8.3	0.085	0.016	5.5*	< 0.001	0.027	0.031	0.8	> 0.1
800–1000 ms	24.6	26.6	24.7	20.9	2.8	0.025	0.015	1.6	> 0.1	0.9	0.011	0.015	0.7	> 0.1	0.056	0.031	1.8(*)	= 0.07
1000–1200 ms	19.8	22.5	24.0	15.0	1.6	0.016	0.014	1.1	> 0.1	3.1	0.031	0.014	2.2*	0.03	0.113	0.029	3.9*	< 0.001

M: Match/no-BEI; M, BEI: Match/BEI; MM: Mismatch/no-BEI; MM, BEI: Mismatch/BEI. * indicates significance level at 0.05 or lower.

the BEI sentences (b = 0.054, SE = 0.021, t = 2.6, p = 0.01), but not in the no-BEI sentences (β < 0.001).

During the time slice from 1000 to 1200 ms post-embedded noun, there was a main effect of BEI and an interaction, but no main effect of classifier incongruity. Follow-up tests showed that the presence of BEI decreased proportion of fixations to the embedded noun picture compared with the absence of BEI only in the classifier-mismatch conditions (by 9%; b = 0.076, SE = 0.019, t = 4.0, p < 0.001), but not in the classifier-match conditions (by −2.7%; p > 0.1).

Proportion of fixations to the target picture

Figure 6 shows the proportion of fixations to the target picture from the onset of the embedded noun across the four conditions. Given that our major interest is to test when the RC parse was built and whether the presence of the embedded NP modulate the quality of RC representations prior to the head noun, we report statistical results all the way up to the end of the sentence, as in Table 3.

As shown in Table 3, we found a significant effect of BEI in four consecutive time slices (200–400 ms, 400–600 ms, 600–800 ms, 800–1000 ms: t's > 2.2). The presence of BEI reliably guided participants to fixate more on the target noun picture than the absence of BEI.

During the time slices of 1000–1200 ms post-embedded noun (when the verb-DE was heard), there were no effects of classifier incongruity or BEI, but a significant interaction. Follow-up tests revealed a marginally significant decrease of looks to the target noun picture in the Mismatch condition than in the Match condition only in no-BEI sentences (by 3.4%, b = 0.043, SE = 0.022, t = 1.9, p = 0.06), but not in the BEI sentences (b = 0.028, SE = 0.022, t = 1.3, p > 0.1).

No other effects were found in the next two time slices (1200–1400 ms, 1400–1600 ms), suggesting that by the time the head noun was processed, participants in all four conditions were equally likely to reliably fixate to the target picture.

Probability of saccades moving into the target picture

Figure 7 shows the proportion of saccades (1) moving away from the embedded NP picture (Figure 7a), (2) moving into the target picture (Figure 7b), (3) moving into the Contrast (Figure 7c) and (4) moving into the Distractor (Figure 7d), with four 400 ms time windows after the onset of the embedded NP. We focused on the statistical analyses up to the disambiguating verb, i.e. 1200 ms after the onset of the embedded noun.

Overall saccades towards the target (regardless of their origin). During the time slice from 0 to 400 ms post-embedded noun, there was a significant main effect of BEI (b = 1.151, SE = 0.478, z = 2.4, p = 0.016), suggesting a larger probability of such saccades in BEI sentences than no-BEI sentences. During the time slice from 400 to 800

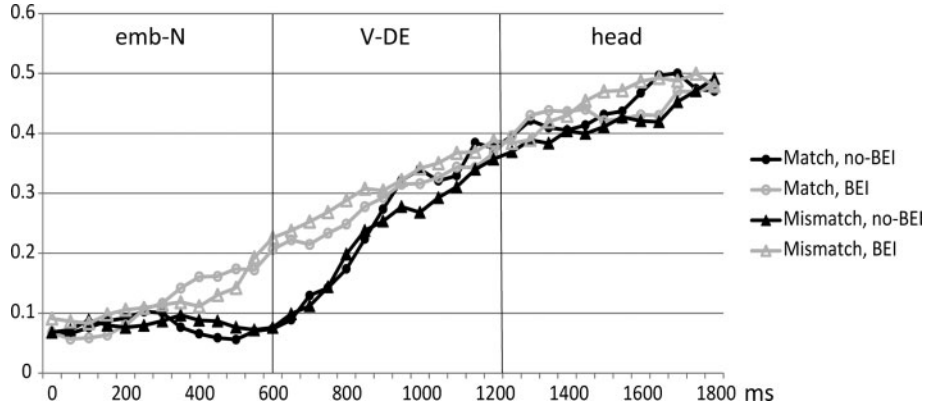

Figure 6. Proportion of fixations to the target picture in the four conditions in Experiment 1 (0 ms marks the average onset of embedded nouns across the four conditions).

Downloaded by [Peking University] at 07:40 19 October 2014

ms post-embedded noun, in addition to the BEI effect consistent with the that of an earlier time window (b = 0.662, SE = 0.317, z = 2.1, p = 0.036), a significant main effect of classifier incongruity was also found (b = 0.628, SE = 0.318, z = 2.0, p = 0.048). This suggests that mismatching classifiers could elicit a stronger prediction for the RC parse, thus inducing more saccades entering the target region as compared with matching classifiers. During the time slice of 800–1200 ms post-embedded noun, there were no main effects of BEI or Classifier but a significant interaction between BEI and Classifier (b = 1.496, SE = 0.638, z = 2.3, p = 0.019). Further analyses showed that mismatching classifiers increased the probability of saccades entering the target region in this time slice only when BEI was absent in the sentences (b = 0.941, SE = 0.535, z = 1.8, p = 0.079), not when BEI was present (p > 0.1).

Saccades specifically from the embedded noun picture to the target. During the time window of 400–800 ms post-embedded noun, there was a (marginally) significant main effect of BEI (b = 0.944, SE = 0.492, z = 1.9, p = 0.055), showing again that participants were discouraged from forming the direct-object parse, but were encouraged to build the RC parse by the syntactic cue. In the time window of 800–1200 ms, similar to the *overall saccades* pattern in this time slice, we only found a significant interaction between the classifier incongruity and the presence of BEI (b = 1.724, SE = 0.896, z = 1.9, p = 0.054), but the simple-effect tests showed that neither the mismatching classifier effect nor the BEI effect was significant (p's > 0.1). In all the time windows, the comparisons between matching and mismatching classifiers were not significant (p's > 0.1).

Discussion

The eye movement data are inconsistent with the DLT. The DLT predicts no effects whatsoever after the embedded NP was presented, but our data show main effects of classifier incongruity, main effects of BEI, and interaction as soon as the embedded NP was heard up until the end of the sentence. Our data better fit with the predictions of the cue-based retrieval theory.

One key finding is that *prior to* the onset of the disambiguating verb (i.e. 200 ms after the onset of the embedded noun), participants could use the retrieval cues provided by the phrasal head of mismatching classifier and BEI to deactivate the incorrect direct-object parse, as indicated by the consistently lower proportion of looks to the embedded noun picture in the Mismatch/no-BEI, Match/BEI, Mismatch/BEI conditions than in the Match/no-BEI condition.

Crucially, when participants were hearing the embedded noun 'football', the significantly lower proportion of looks to the embedded noun picture in the Mismatch/no-BEI condition than in the Match/no-BEI condition shows that participants were sensitive to the classifier-noun incongruity, and were using the mismatching classifier cue to deactivate the incorrect direct-object parse. The aural presentation mode in our experiment renders the classifier-noun incongruity apparent to native speakers' ears, making it unlikely for participants to treat the incongruent classifiers as 'typos' (speech errors), or incorrectly attaching the local noun 'football' to the preceding mismatching classifier *shan* (cf. Hsu et al., 2006).

Perhaps more convincing evidence for the cue-based retrieval theory comes from the reliably more proportion of looks as well as larger probability of saccades to the target noun picture in the two BEI sentences compared with the two no-BEI sentences, shortly after the onset of the embedded noun and before the offset of the verb. This confirms that participants were able to use the syntactic BEI cue to construct the RC parse and to identify the correct picture, as well as to reduce dwelling on the non-target pictures.

We found no evidence from *the proportion of fixations* that participants were cued by the mismatching classifier

Downloaded by [Peking University] at 07:40 19 October 2014

Table 3. Linear mixed-effects regression model for the proportion of looks to the target after the onsets of embedded noun across the four conditions (Exp. 1)

					CL incongruity					BEI					Interaction			
Time windows	M (%)	M, BEI (%)	MM (%)	MM, BEI (%)	Mean difference (%)	β	SE	t	p	Mean difference (%)	β	SE	t	p	β	SE	t	p
20–400 ms	9.1	11.0	8.5	11.0	0.4	0.001	0.011	0.1	> 0.1	2.2	0.024	0.011	2.2*	0.03	0.010	0.022	0.4	> 0.1
400–600 ms	6.2	16.5	8.1	14.3	0.2	0.001	0.011	0.1	> 0.1	8.2	0.085	0.011	7.6*	< 0.001	0.038	0.022	1.7	0.09
600–800 ms	10.9	21.6	10.8	24.4	1.3	0.014	0.013	1.1	> 0.1	12.2	0.128	0.013	9.8*	< 0.001	0.025	0.026	1.0	> 0.1
800–1000 ms	25.1	28.3	24.6	30.3	0.8	0.01	0.016	0.7	> 0.1	4.5	0.051	0.016	3.3*	0.001	0.028	0.031	0.9	> 0.1
1000–1200 ms	34.4	33.3	31.0	35.4	0.6	0.009	0.016	0.5	> 0.1	1.7	0.021	0.016	1.3	> 0.1	0.069	0.033	2.1*	0.03

M: Match/no-BEI; M, BEI: Match/BEI; MM: Mismatch/no-BEI; MM, BEI: Mismatch/BEI. * indicates significance level at 0.05 or lower.

to dwell more on the target, as shown by the lack of overall significant difference between the two no-BEI sentences in terms of the proportion of looks to the target noun picture. We also found no effect of classifier incongruity by the measure of saccades launched specifically from the embedded noun picture towards the target picture. However, evidence from the probability of *overall saccades towards the target* indicates that this lexical cue *was* utilised by participants to anticipate the RC structure, as more saccades were redirected to the target picture in sentences with a mismatching classifier than those with a matching classifier. Thus, taken together, our findings seem to show that the mismatching classifiers may help participants to avoid the incorrect direct-object parse and, to a certain extent, to build the RC parse. The RC-prediction effect is clearly more subtle compared with the strong effect of syntactic cue BEI.

It is worth noting that indeed the intervening embedded NP interfered with repeated retrieval of the head NP in the Mismatch/no-BEI condition, as reflected by fewer looks to the target picture in this condition than in the Match/no-BEI condition at the 1000–1200 ms time window when the disambiguating verb was heard. As predicted by the cue-based retrieval theory, such interference effects are likely to occur during long-distance integrations of a mismatching classifier to its host noun (i.e. the head of RC) and of the head NP as an argument to its matrix verb. We suggest that such interference effects in processing non-local dependencies might be augmented by the visual world paradigm.

Motivations for Experiment 2

Our visual world eye-tracking experiment found reliable evidence for highly incremental parsing from the passive marker BEI, but not that consistent from the mismatching classifier. It is plausible that the statistical analyses of our eye-tracking data were not sensitive enough to thoroughly reveal effects of classifier incongruity, due to the limitation of existing statistical resolutions for data from visual world paradigm, particularly in the presence of more than two dependent variables (i.e. four candidate pictures to choose from as in our study) (Barr, 2008). Thus, it is worthwhile to replicate the findings using a different paradigm, for instance, self-paced reading.

Furthermore, an alternative explanation for the patterns of fixation (and of saccades) is due to the ambiguous status of the embedded noun in the Match/no-BEI condition, which forced syntactic reanalysis, whereas no such reanalysis was involved in the other conditions. This possibility could be evaluated in a self-paced reading experiment with punctuation marks displayed, which may give additional hints to comprehenders about whether the encountered noun is the end of the sentence, or is followed by other constituents that are highly likely to belong to an RC

Table 4. Linear mixed-effects models for log-RTs by regions in Experiment 2.

Region	Predictors in final model	β	SE	t	p
Embedded noun	Intercept (cl/bei)	6.313	0.043	146.01	<0.001
	cl = mismatch	−0.029	0.026	−1.12	0.26
	bei = no-BEI	0.015	0.026	0.57	0.57
	cl x bei	0.042	0.037	1.14	0.25
V-DE	Intercept (cl/bei)	6.251	0.029	212.36	<0.001
	cl = mismatch	0.014	0.026	0.56	0.58
	bei = no-BEI	0.122	0.031	3.98	<0.001
	cl x bei	−0.009	0.037	−0.24	0.807
Adjective	Intercept (cl/bei)	6.303	0.032	194.44	<.0001
	cl = mismatch	0.011	0.031	0.35	0.73
	bei = no-BEI	0.089	0.025	3.53	<0.001
	cl x bei	−0.072	0.036	−2.01	0.044
Head Noun	Intercept (cl/bei)	6.456	0.04	160.19	<.0001
	cl = mismatch	−0.052	0.037	−1.42	0.16
	bei = no-BEI	0.029	0.029	1.00	0.32
	cl x bei	0.001	0.042	0.03	0.98

Downloaded by [Peking University] at 07:40 19 October 2014

structure. In other words, we were motivated to investigate the processing consequences of the pre-RC cues under reduced garden-path circumstances.

Experiment 2: self-paced reading

Experiment 2 aimed to replicate the major findings in Experiment 1 using the self-paced reading paradigm, and to evaluate the predictions made by the two theories under mild garden-path circumstances in which punctuation marks are inserted.

The stimuli including fillers were maximally similar to those in Experiment 1, except two modifications. First, in setting up a 2-referent context (e.g. 'two windows'), we added modifying adjectives (e.g. colour, shape, style, positions, ownership, status, materials, etc.) to distinguish

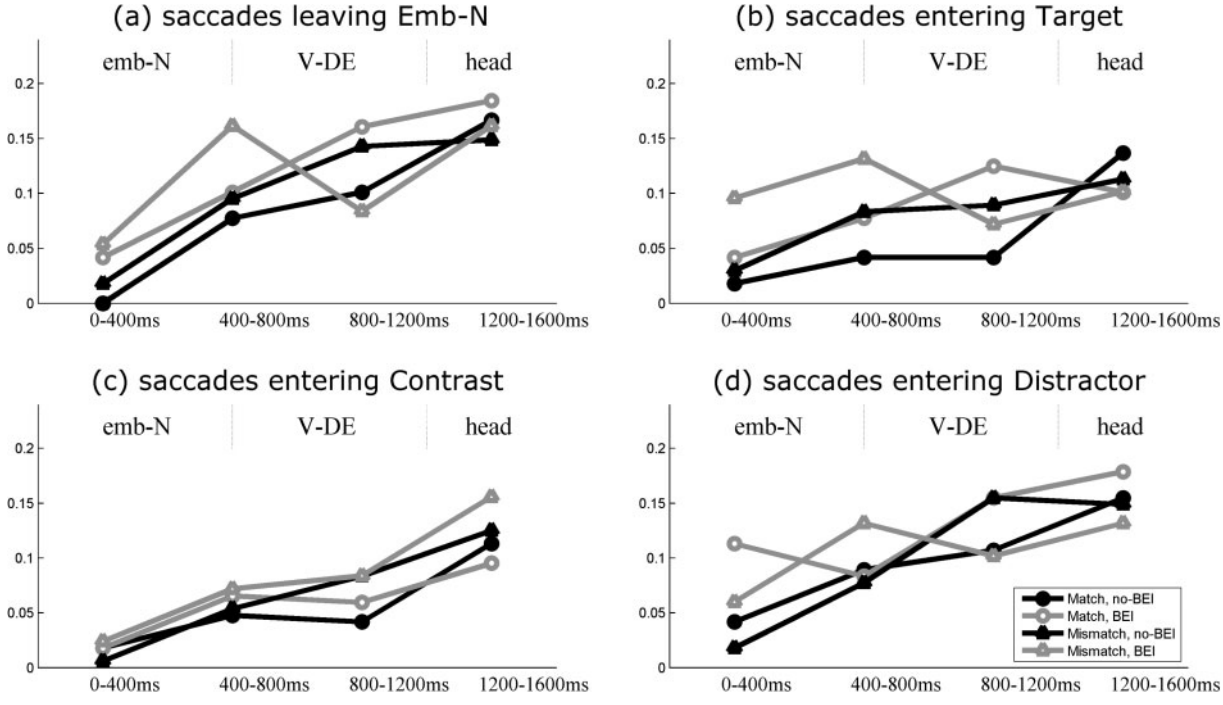

Figure 7. Proportion of saccades that were (a) leaving the embedded noun picture (on the top left), (b) entering the target picture (on the top right), (c) entering the contrast picture (on the bottom left) and (d) entering the distractor picture (on the bottom right) in the four conditions in Experiment 1.

Downloaded by [Peking University] at 07:40 19 October 2014

the two entities ('the French window' vs. 'the inlaid window'). Second, we further extended the structure of the critical stimuli by introducing a female character Xiaomei (or a male character Xiaoming) serving as the matrix subject, using the format of 'S ADV V [$_{RC}$ …] ADJ O'. A sample set of Mismatch trials (with or without BEI) is shown in (10). The context sentences were presented sentence-by-sentence as indicated by double slashes (//), and the critical sentence was presented phrase-by-phrase as indicated by slashes (/).

(10) Sample Mismatch trials (with or without BEI):
Qiangmianshang you liang-shan chuanghu (on-the-wall have two-CL window '*There are two windows on the wall*'). //

Yizhi zuqiu dasuile luodishi bolichuang (one-CL football break-ASP French-style window '*A football broke the French-style window*'). //

Yikuai zhuantou dasuile neiqianshi bolichuang (one-CL brick break-ASP inlaid window '*A brick broke the inlaid window*'). //

Target sentence:

Xiaomei / xiaoxinde / yikai / na-shan / (bei) / zuqiu / dasuide / luodishi /chuanghu.//

Xiaomei/ carefully / move-away/ that-CL$_{*football/window}$/ (PASS) / football/ broke-DE / French / window.//

'Xiaomei carefully moved away the French window that {the football broke | was broken by the football}'.

The Match trials had exactly the same format, except that the classifier *zhi* matched the embedded noun *zuqiu* ('football') as well as the head noun of the RC (*huaping* 'vase'), as in (11).

(11) Sample Match trials (with or without BEI):

Zhuoshang fangzhe liang-zhi huaping (on-the-table lay-ASP two-CL vase '*There are two vases on the table*').//

Yizhi zuqiu dasuile lansede huaping (one-CL football break-ASP blue vase '*A football broke the blue vase*'). //

Yikuai zhuantou dasuile hongsede huaping (one-CL brick break-ASP red vase '*A brick broke the red vase*'). //

Target sentence:

Xiaomei / xiaoxinde / yikai / na-zhi / (bei) / zuqiu / dasuide / lansede / huaping.//

Xiaomei/ carefully / move-away/ that-CL$_{football/vase}$/ (PASS) / football/ broke-DE /blue /vase. //

'Xiaomei carefully moved away the blue vase that {the football broke | was broken by the football}'.

One consequence of using the self-paced reading paradigm is the (partial) removal of ambiguity by the absence of a punctuation mark after the embedded NP. Normally in presenting words in self-paced reading, the last word of the sentence is accompanied by an immediately following sentence-ending period. Given that our stimuli (both target and filler items) consisted of either RCs or simple clauses (i.e. the direct-object parse), over time participants might learn to take the absence of a period after the embedded NP as a signal for a complex structure.[5]

With the RC-biasing contexts and the additional disambiguating cue of (the absence of) punctuation marks, the RC parse should be highly expected in all four conditions. Thus, the DLT predicts no processing differences at the embedded NP and all the way up to the end of the sentence (cf. predictions outlined in Table 1). In contrast, the cue-based retrieval theory predicts processing differences across conditions reflected by a main effect of BEI and an interaction between Classifier and BEI, possibly beginning as early as the embedded noun region and continuing to the head noun where various dependencies/attachments are established.

Method

Participants

Eighty native speakers of Mandarin Chinese participated in the experiment in exchange for 15 RMB. All were undergraduate students from Shanghai International Studies University. Their mean age was 21.85.

Design and materials

Twenty-four sets of experimental items were constructed following the format of (10–11), each consisting of four sentences. In the critical sentence, the semantic congruity of classifiers (match or mismatch) and the passive marker BEI (presence or absence) were crossed, resulting in four conditions: Match/no-BEI, Match/BEI, Mismatch/no-BEI, and Mismatch/BEI. The head nouns in the four conditions were matched in terms of lexical (log) frequency, visual complexity (in terms of number of strokes) and word length.

As in Experiment 1, there were 36 filler items in addition to the experimental items. In the critical sentence, 20 fillers contained a simple noun phrase as in (12), 6 fillers contained a subject-extracted RC as in (13a) and 10 fillers contained had the structure as in (13b)

(12) Sample Filler Item: direct object noun phrase
'There were two watermelons in the zoo. // A zebra ate the seedless watermelon. // A rhinoceros ate the black-seed watermelon. // Xiaomei looked in surprise at **the zebra**'.

(13) Sample Filler Item:
a. Subject-extracted RC
'There were two helicopters on the runways. // The private helicopter flew over a snow mountain. // The military helicopter flew over a desert. // Xiaomei proudly piloted **the private helicopter that flew over the snowy mountain**'.

b. Direct object noun phrase
'There are two fishing rods by the pond. // The electric fishing rod has a caterpillar on it. // The bamboo-made fishing rod has an earthworm on it. // Xiaoming naughtily plays with **the earthworm**.

Downloaded by [Peking University] at 07:40 19 October 2014

Thus, of all the 60 sentences participants read, half contained an RC (12 ORCs and 18 SRCs including 12 passive RCs), and half contained a simple noun phrase.

Four stimulus lists were created using a Latin Square design. Each list contained one version of each experimental item, along with all 36 of the filler items, for a total of 60 items. The order of items was pseudo-randomised for each list, such that all target items were separated by at least one filler, and no target items occurred in the first two trials for each list.

Procedure

A word-by-word moving-window self-paced reading experiment was run on a PC with a 21-inch CRT monitor using Linger software developed by Doug Rohde. Each trial began with a series of dashes marking the position and length of the words in the sentences, across several lines of text. Participants pressed the spacebar to reveal each fragment of the materials. The first three sentences were presented sentence-by-sentence. The final critical sentence was presented word-by-word. Participants were instructed to read the sentences for comprehension. RTs for each word were recorded. For each trial, participants read the sentences at their own speed, and then answered a yes/no comprehension question.

The experiment was preceded by six practice trials. All items were followed by yes/no comprehension questions. The questions asked about different parts of the sentences in order to encourage the participants to focus equally on all parts of each sentence. Half of the comprehension questions had 'yes' answers, the other half had 'no' answers. Participants answered the questions by pushing the F key for 'yes' and the J key for 'no'. The computer flashed 'You are wrong' in Chinese if the questions were incorrectly answered, but no feedback was provided if the answers were correct.

Results

Eleven participants' results were omitted from analyses because of relatively poor comprehension question performance on the target items (accuracy rate ranges from 58.33% to 79.17%). We also analysed all participants' data, which yielded a similar pattern of results.

Question-answering accuracy

Averaging across all target and filler trials, 69 participants answered 91% of the comprehension questions correctly. On the critical trials, the overall accuracy rate across participants was 90%, and all participants answered at least 83.3% of the questions correctly. On filler trials, the overall accuracy rate was 92%.

Reading times

Statistical analyses were conducted by linear mixed-effects modelling with the lme4 package for the statistical language R (version 2.13.0; CRAN project; the R Foundation for Statistical Computing, 2011). To choose final best-fitting models for each region, we included all main effects and interactions in the fixed and random effects, and then simplified the random-effects models stepwise until the credibility of any smaller model having generated the data in comparison to the fuller model was below $p = 0.05$ by the log-likelihood χ^2 test. Significant p-values were estimated by means of the function pvals.fnc from the language package using Markov Chain Monte Carlo simulation for final models with random intercepts only, or from the t distribution using degrees of freedom estimated by the number of observations minus the number of fixed effects for final models with random slopes and intercepts (Baayen et al., 2008).

RTs longer than 5000 ms were excluded from further analysis (15 data points). Then the remaining individual RTs were log-transformed[6] to correct for the heavily skewed distribution. Using Turkey's fences, we removed outliers that were above the third quartile plus 1.5 times the interquartile range for each condition within each word region. This affected 3.57% of the total data. After elimination of outliers, separate mixed-models of log RT were fitted to each word region after BEI (totaling four positions), following the stepwise model simplification procedure described above until the final, best fitting model was obtained. Figure 8 presents the mean reading times for all word positions. Table 4 presents the statistical results for each region.

At the embedded noun ('football'), there were no main effects and no interaction, although the BEI conditions were read numerically faster (by 77 ms) than the no-BEI conditions.

At the verb-DE region, there was a significant main effect of BEI ($t = 3.98$, $p < 0.001$). Conditions with BEI were read 194 ms faster than conditions without BEI. There was no main effect of classifier incongruity and no interaction.

At the adjective region, there was a main effect of BEI ($t = 3.53$, $p < 0.001$) and an interaction between classifier incongruity and BEI ($t = -2.01$, $p < 0.05$), but no main effect of classifier incongruity ($t = 0.35$). Planned comparisons showed that the presence of BEI facilitated RTs in the match conditions (−92 ms; $b = -0.089$, SE = 0.027, $t = -3.28$, $p = 0.001$), but not in the mismatch conditions (17 ms; $t = -0.67$). In addition, the presence of mismatching classifier facilitated RTs in the no-BEI conditions (−65 ms, $b = -0.06$, SE = 0.027, $t = -2.14$, $p = 0.033$), but not in the BEI conditions (10 ms, $b = 0.014$, SE = 0.023, $t = 0.62$).

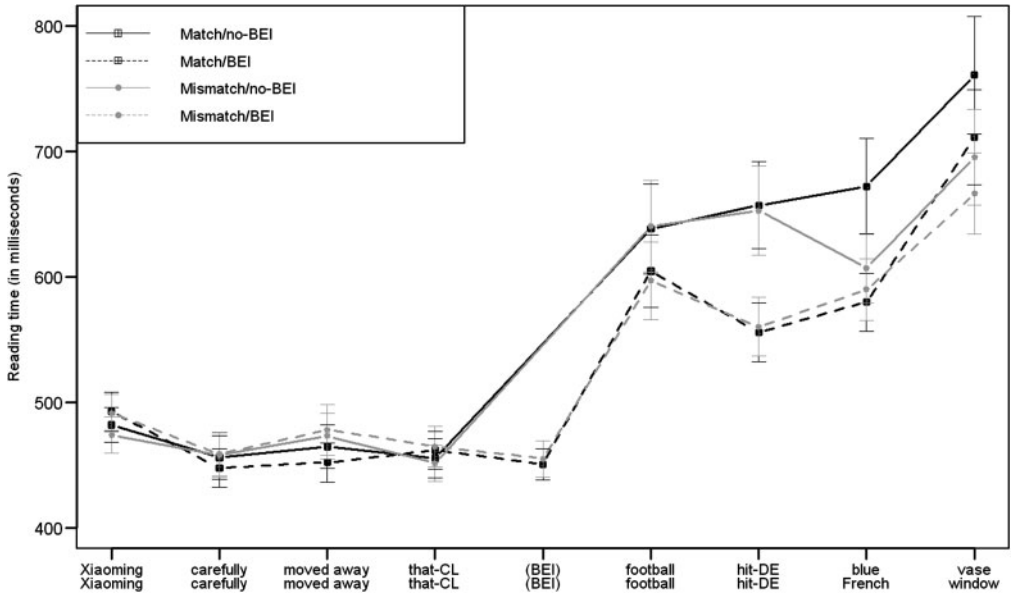

Figure 8. Reading times per word across the four conditions in the critical sentence in Experiment 2.

At the sentence-final head noun region, there were no main effects and no interactions.

Discussion

The RT evidence in Experiment 2 is generally consistent with the findings in Experiment 1. Contrary to what the DLT predicts, processing differences were found at the verb-DE and Adjective regions even when ambiguity was virtually 'eliminated' in the Match/no-BEI condition. The results of Experiment 2 are consistent with the predictions of the cue-based retrieval theory.

The faster RTs in the BEI conditions than in the no-BEI conditions at the verb-DE region clearly suggest that the BEI cue was effective in helping participants to build RC structures. The interaction at the adjective region was mainly due to the slowest RT in the Match/no-BEI condition, with the other three conditions being processed about equally fast. This suggests that beyond the disambiguating point (i.e. the verb) or at the spillover region of the adjective, both BEI and mismatching classifier are reliable cues for RCs, but their effects are not additive. Note however, the BEI cue did have an early effect, given that the effect of BEI occurred at the disambiguating verb and the effect of classifier (as reflected by the interaction) did not appear until the spillover region. This pattern is highly consistent with our saccades data in Experiment 1.

The absence of punctuation marks after the embedded NP appears to have not only greatly reduced the potential structural ambiguity in the Match/no-BEI condition at the embedded noun, but also much weakened the potential lexical disruption induced by the classifier-noun incongruity in the Mismatch/no-BEI condition commonly observed in previous work (cf. Hsu, 2006; Hsu et al., 2006). This is reflected by the similar RTs between the two no-BEI conditions at the embedded NP and verb-DE regions, even though differences *were* observed at later words. We speculate that the lack of lexical disruption effect in our experiment might be due to a highly incremental parser being saturated by richly available information – the RC-biasing context, the presence of mismatching classifier, and the lack of period – to determine that the final parse is the RC. Thus the lexical disruption effects in the Mismatch/no-BEI condition, if any, are neutralised with the reanalysis process in the Match/no-BEI condition, given that the reanalysis is rendered much easier by the punctuation cue.

It is worth noting that given that the two passive RC sentences are grammatically subject-extracted RCs, the main effect of BEI at the verb-DE region may in part be attributable to the universal subject preference (Keenan & Comrie, 1977). We believe this explanation cannot be the whole story, because otherwise one shall expect to see no facilitative effects in the two (no-BEI) ORC conditions. But such effects were observed in the Mismatch/no-BEI condition, as shown by the significant interaction between classifier and BEI at the Adjective, where the Mismatch/no-BEI ORC sentences were read faster than the Match/no-BEI ORC sentences. Interestingly towards the end of the sentence, the Match/no-BEI condition was read slowest. We take this as evidence that in the absence of cues, Chinese ORCs are intrinsically difficult to parse.

Downloaded by [Peking University] at 07:40 19 October 2014

Downloaded by [Peking University] at 07:40 19 October 2014

General discussion

We presented two experiments that investigated the processing of object-modifying RCs in Chinese by manipulating the congruency between classifiers and nouns and the presence of the passive marker BEI between verb-argument dependencies in object-modifying RCs in Mandarin, and compared the predictions in these cases made by two working-memory-based theories, namely the DLT and the cue-based retrieval theory. By creating 2-referent discourse contexts prior to the critical RC sentences, we obtained converging evidence from our visual world eye-tracking and self-paced reading experiments, demonstrating that Chinese comprehenders were sensitive to the availability of pre-RC heads (i.e. mismatching classifier alone or in conjunction with BEI) as cues to deactivate the direct-object parse as soon as the embedded noun was presented, and pre-build the head-final RC structure before the head noun was shown in the sentence-final position. The results were consistent with the predictions of the cue-based retrieval theory, but contrary to the predictions of the DLT.

As discussed in the prediction section of Experiment 1, the integration-cost metric of the DLT predicts no processing differences across our experimental conditions, simply because all intervening discourse referents in various dependencies were introduced earlier in the preceding contexts, thus involving no extra cost. The storage-cost metric of the DLT predicts that processing differences should arise prior to the disambiguating RC-internal verb, but no other effects are expected in regions beginning from the verb. However, both our eye-tracking data and RT data clearly revealed effects of classifier-incongruity, of BEI, and of classifier interacting with BEI beginning at the disambiguating verb and continuing till the end of the sentence. All this imposes challenges to the storage-cost metric of the DLT.

Rather, these effects provide supporting evidence for the cue-based retrieval theory, which predicts that if pre-head constituents (or cues) that occur early in head-dependent relationships help pre-activate an upcoming head, then accessing that head becomes easier than if no such cues are available to predict it. Specifically, in our experimental setting, given the preceding discourse contexts, the cue-based retrieval theory predicts that the mismatching classifier alone and in conjunction with the passive marker BEI should help Chinese comprehenders not to entertain the direct-object parse as soon as the embedded noun is available, and depending on their cueing strength to activate the head noun, mismatching classifier interacts with BEI to build the correct RC parse well before the head noun is presented.

Regarding the strength of BEI as a syntactic, unambiguous retrieval cue for the upcoming head of RC with little interference from the intervening embedded noun, both our RT data and eye movement data are consistent with the predictions of the cue-based retrieval theory. In Experiment 1, we found a reliably greater proportion of looks to the target noun picture and less proportion of looks to the embedded noun picture in sentences with BEI compared with sentences without BEI, shortly after the onset of the embedded noun and before the offset of the verb. This effect of BEI was further confirmed by our saccades data showing that the decreased proportion of looks to the embedded noun picture was due to fixations redirected to the target picture, leading to the increased proportion of looks to the target picture. In Experiment 2, we found the facilitative effects of BEI at the verb and adjective, as reflected by significantly faster RTs in sentences with BEI than in sentences without BEI. These results confirm that our Chinese participants utilised the BEI cue to successfully retrieve the head noun, thereby constructing the RC parse efficiently. Our finding for BEI as a strong predictor for RC is also consistent with the findings in Wu, Kaiser, and Andersen (2009) and Wu and Kaiser (submitted), but our study is perhaps the first one that uses object-modifying RCs to verify this.

The role of the mismatching classifier cue was also found to be in line with the prediction of the cue-based retrieval theory in both experiments. Mismatching classifiers succeeded in helping our participants to deactivate the incorrect direct-object parse, but their impact on predicting the correct RC structure is most likely subtle and delayed. In Experiment 1, we found a significantly lower proportion of looks to the embedded noun picture in sentences with mismatching classifiers compared with sentences with matching classifiers, suggesting that participants were using the classifier mismatch cue not to look at the 'wrong' picture (i.e. the embedded noun). Facilitations for RC-structure building were observed in terms of the probability of overall saccades towards the target picture between sentences without BEI, although such facilitations were not detected using other measurements. In Experiment 2, we found a delayed effect of classifier-noun incongruity at the adjective region for no-BEI sentences: the Mismatch/no-BEI condition was read faster than the Match/no-BEI condition. The delayed facilitative effects suggest that the mismatching classifier is a valid (though subtle) cue for constructing RC structure.

Taken together, our data jointly suggest that as cues for anticipating and constructing the RC structure, the syntactic marker BEI carries more weight than the lexical cue of mismatching classifier in driving retrieval of the head noun and in effectively limiting interference from the embedded noun. Nevertheless, these two differently weighted cues work together to help Chinese comprehenders to build the RC structure incrementally and probabilistically, as evidenced by the significant Classifier*BEI interactions found in both experiments. In Experiment 1, we found that the presence of the BEI cue significantly

Downloaded by [Peking University] at 07:40 19 October 2014

lowered proportion of looks to the embedded noun picture than the absence of BEI only in sentences with mismatching classifiers, both in an early time window (i.e. 200–400 ms after the onset of the embedded noun) and a late time window (i.e. 800–1000 ms after the onset of the embedded noun). The data of probability of saccades, on the other hand, showed that comprehenders could redirect fixation to the target picture at the earliest time window when provided with double cues (i.e. Mismatch/BEI), and at relatively earlier time windows when provided with one cue (i.e. Match/BEI; Mismatch/no-BEI) than no cues (i.e. Match/no-BEI). In Experiment 2 at the adjective region, the Match/no-BEI condition was read the slowest among the other conditions in which cues are present. All this suggests that both types of cues are reliable in helping build the correct RC parse, although the timing for them to take effects varies, and their effects may not be additive beyond the disambiguating point.

To conclude, the results of two experiments using Chinese object-modifying RCs preceded by discourse contexts help evaluate the two memory-based theories of sentence processing. We show that contrary to what the DLT would predict, interposing pre-RC heads of classifiers and/or BEI in dependency relationship does not necessarily mean increased processing difficulty. Rather, as the cue-based retrieval theory predicts, those incomplete heads yet to be integrated into various dependency relationships may help pre-build complex RC structure, and the facilitations of RC processing may vary depending on the strength of cues.

Acknowledgements

We would like to thank audiences at the Workshop on Typology, Acquisition, & Processing of Relative Clauses (Leipzig, Germany, June 2005) and the 19th Annual CUNY Conference on Human Sentence Processing (New York, March 2006), where earlier versions of this line of research were presented. The stimuli used in this research are a revised version of the stimuli used in Wu (2009). We gratefully acknowledge Todd Haskell for his indispensable contributions to earlier incarnations of the visual world eye-tracking research, and for his useful comments on an early version of this paper. We thank Yanan Sheng for assistance in running the self-paced reading experiment, Yunyan Duan for helping analyse and plot the eye-tracking data, Ming Yan, Junru Wu and Xiaoming Jiang for help with the eye-tracking experiment, and Steve Politzer-Ahles for helpful discussion. Thanks also go to Elaine Andersen and Elsi Kaiser for feedback on earlier stages of this research. This research was supported by grants from Shanghai Pujing Project (10PJC066), Shanghai Municipal Philosophy and Social Sciences Foundation (2010BYY003) and National Social Science Foundation of China (13BYY152) to the first author, and by a grant from National Social Science Foundation of China (12&ZD119) to the third author.

Notes

1. Recent developments of DLT seem to vary in their specifications of processes involved in sentence comprehension. Gibson and Wu (2013) suggest (1) storage costs and (2) retrieval/integration costs, without distinguishing retrieval and integration costs. Levy, Fedorenko, and Gibson (2013) suggest (1) storage and retrieval costs and (2) integration costs. In this paper, we focus on the early version of DLT.
2. Here we use RC without specifying its extraction type. According to generative approaches to syntactic analyses of passivisation (Chomsky, 1981; Huang, Li, & Li, 2008), RCs with the presence of BEI are subject-extracted RCs (SRCs), because syntactically the head noun 'book' is base-generated in the subject position, as in the English passive 'The book is criticised by the professor'. Semantically, it is still the theme, serving as the complement of the verb 'criticise', regardless of whether it is the subject or object.
3. It is not straightforward to tease apart the predictions in terms of the eye movement pattern by the two theories when the embedded noun is presented. The storage-cost metric of the DLT predicts easier processing in the Match/no-BEI condition (where less heads are projected) compared with the other three conditions (where more heads are projected). This processing ease may be reflected by reliable looks to the embedded noun picture in the Match/no-BEI condition, because the structurally simpler direct-object parse is preferably entertained or highly activated than the RC parse. But it is not clear what kind of eye gaze pattern might be induced by processing difficulty in the other three conditions. The Cue-based Retrieval Theory also predicts prolonged looks to the embedded noun picture in the Match/no-BEI condition compared with the other three conditions where either one cue (i.e., mismatching classifier) or two cues (i.e. mismatching classifier and BEI) will help participants not to entertain the incorrect direct-object parse. Thus, we refrain from further discussing predictions when the embedded noun is presented, but focus on contrastive predictions that the two theories make.
4. An anonymous reviewer suggested us to run a statistical analysis, comparing saccades from the embedded noun picture *to the target picture* with saccades from the embedded noun picture *to other non-target pictures*. If we found main effects of BEI and classifier for entries to the target picture (and the lack of such main effects for entries to the distractor and contrast pictures), then we could safely say that more looks to the target picture was a result of escape from the embedded NP picture, showing that the presence of a cue discourages a direct-object parse and facilitates an RC parse. Although we sympathise with this idea, we believe that this analysis is not appropriate, for the following reasons: First, there is a high risk of a type I error due to multiple comparisons when simultaneously applying generalised linear mixed models with binominal family to multiple contrasts of saccades. If the saccades are categorised into more than two levels, i.e., saccades towards *target*, *distractor*, *contrast* and the blank space, it is inappropriate to apply either logistic regression or generalised linear mixed model with logit link function. Second, even if we run the log-linear regression analysis which is suitable for analysing multinominal data, that model cannot incorporate the estimates of random effects, which, in our case, means that it has to ignore the variance between participants and between items.
5. We note that the probabilities are rather low for another word to follow the 'football' in the main-clause analysis. In Chinese, temporal adverbs like 'yesterday' or locative adverbial phrase like 'in the garden' cannot occur after the object noun phrase 'football'. Also, given the highly constraining discourse/sentential contexts (e.g., the verb

'move away' and the noun 'football' without any modifications), DE plus another noun (e.g., 'football's dust') is quite unlikely due to pragmatic oddity.

6. In addition to analysing log RTs, we also analysed negative reciprocal transformed RTs with or without the trimming procedure of Turkey's Fences. All of our analyses resulted in the same statistical patterns. Although the box-cox function (Kliegl, Masson, & Richter, 2010) suggests that a negative reciprocal transformation of RT (i.e., −1/RT) best fits our reading time data, extremely small values of reciprocal RTs also masked the outliers in the form of raw RTs, making it difficult to interpret statistical results against the graph based on raw RT data. We therefore report the log RT data.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238. doi:10.1016/0010-0277(88)90020-0

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. doi:10.1016/j.jml.2007.12.005

Barr, D. J. (2008). Analysing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457–474. doi:10.1016/j.jml.2007.09.002

Bartek, B., Lewis, R., Vasishth, S., & Smith, M. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1178–1198. doi:10.1037/a0024194

Chomsky, N. (1981). *Lectures on government and binding*. *Studies in Generative Grammar*9. Dordrecht: Foris.

Clifton, C., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In G. Carlson & M. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 273–217). The Netherlands: Springer.

De Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America*, *96*, 2037–2047. doi:10.1121/1.410145

Dryer, M. S. (1992). The Greenbergian word order correlations. *Language*, *68*, 81–138. Retrieved from http://www.jstor.org/stable/416370

Fedorenko, E., Piantadosi, S., & Gibson, E. (2012). Processing relative clauses in supportive contexts. *Cognitive Science*, *36*, 471–497. doi:10.1111/j.1551-6709.2011.01217.x

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading*. Hillsdale, NJ: Erlbaum.

Frazier, L., & d'Arcais, F. (1989). Filler-driven parsing: a study of gap filling in Dutch. *Journal of Memory and Language*, *28*, 331–344. doi:10.1016/0749-596X(89)90037-5

Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, *58*, 161–187. doi:10.1016/j.jml.2007.07.004

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

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, and W. O'Neil (Eds.), *Image, language, brain* (pp. 313–353). Cambridge, MA: MIT Press.

Gibson, E., & Pearlmutter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Science*, *2*, 262–268. doi:10.1016/S1364-6613(98)01187-5

Gibson, E., & Wu, H.-H. I. (2013). Processing Chinese relative clauses in context. *Language and Cognitive Processes*, *28*, 125–155. doi:10.1080/01690965.2010.536656

Greenberg, J. H. (1963). Some universals of language with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of human language* (pp. 73–113). Cambridge, MA: MIT Press.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science*, *29*, 261–290. doi:10.1207/s15516709cog0000_7

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10.1–10.112). New York, NY: Wiley.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, 26, 376–405. doi:10.1080/01690965.2010.492642

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

Hsu, C.-C. N. (2006). *Issues in head-final relative clauses in Chinese – Derivation, processing and acquisition*(PhD thesis). University of Delaware.

Hsu, C.-C. N., Hurewitz, F, & Phillips, C. (2006, March). *Contextual and syntactic cues for head-final relative clauses in Chinese*. Paper presented in the 19 th Annual CUNY Conference on Human Sentence Processing, March 23–25. CUNY Graduate Center, New York City.

Hsu, C. -C. N., Phillips, C., & Yoshida, S. (2005). *Cues for head-final relative clauses in Chinese*. Poster presented at the 18th Annual CUNY Sentence Processing Conference. March 31–April 2, Tucson, AZ.

Huang, C. J., Li, A., & Li, Y. (2008). *The syntax of Chinese*. Cambridge: Cambridge University Press.

Lee, T. H.-T. (2006). A note on garden path sentences in Chinese. In D.-A. Ho, S. Cheung, W. Pan, & F. Wu (Eds.), *Linguistic studies in Chinese and neighboring languages: Festschrift in honor of Professor Pang-Hsin Ting on his seventieth birthday* (pp. 491–518). Taipei: Institute of Linguistics, Academia Sinica.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177. doi:10.1016/j.cognition.2007.05.006

Levy, R., Fedorenko, E., & Gibson, E. (2013). The syntactic complexity of Russian relative clauses. *Journal of Memory and Language*. doi:10.1016/j.jml.2012.10.005

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

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in cognitive sciences*, *10*, 447–454. doi:10.1016/j.tics.2006.08.007

Li, W., & Yang, Y. (2009). Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences. *Neuroscience*, *158*, 1416–1425. doi:10.1016/j.neuroscience.2008.10.065

Lin, C. J. C. (2006). *Relative-clause processing in typologically distinct languages: A universal parsing account* (Unpublished PhD dissertation). University of Arizona, Tucson.

Downloaded by [Peking University] at 07:40 19 October 2014

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and language*, *49*, 133–156. doi:10.1016/S0749-596X(03)00023-8

Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic inquiry*, *8*, 63–99.

Kliegl, R., Masson, M.E.J., & Richter, E.M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition*, *18*, 655–681. doi:10.1080/13506280902986058

Kliegl, R., Risse, S., & Laubrock, J. (2007). Preview benefit and parafoveal-on-foveal effects from word n + 2. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1250–1255. doi:10.1037/0096-1523.33.5.1250

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*(1), 15–47. doi:10.1016/0010-0277(72)90028-5

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*, 627–645. doi:10.1023/A:1026528912821

MacWhinney, B., & Pléh, C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition*, *29*(2), 95–141. doi:10.1016/0010-0277(88)90034-0

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information processing time with and without saccades. *Perception and Psychophysics*, *53*, 372–380. doi:10.3758/BF03206780

Nakatani, K., & Gibson, E. (2008). Distinguishing theories of syntactic expectation cost in sentence comprehension: Evidence from Japanese. *Linguistics*, *46*(1), 63–86. doi:10.1515/LING.2008.003

Nakatani, K., & Gibson, E. (2010). An on-line study of Japanese nesting complexity. *Cognitive Science*, *34*(1), 94–112. doi:10.1111/j.1551-6709.2009.01067.x

Phillips, C., & Lau, E. (2004). Foundational issues. *Journal of Linguistic*, *40*, 571–591. doi:10.1017/S0022226704002774

R Core Development Team. (2008). R: A Language and Environment for Statistical computing [computer Software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved April 20, 2008, from http://www.R-project.org

Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either…or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 425–436. doi:10.1037/0278-7393.32.2.425

Steeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *Journal of Acoustic Society of America*, *64*, 1582–1592. doi:10.1121/1.382142

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634. doi:10.1126/science.7777863

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (vol. xviii, pp. 217–262). San Diego, CA: Academic Press.

Tanenhaus, M. K., & Trueswell, J. C. (2006). Eye movements and spoken language comprehension. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 863–900). New York: Academic Press, Elsevier.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influence on syntactic processing: Use of thematic information in syntactic disambiguation. *Journal of Memory and Language*, *33*, 285–318. doi:10.1006/jmla.1994.1014

Van Dyke, J. A., & Lewis, R. (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 Languag*e, *49*, 285–316.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*, 247–263. doi:10.1016/j.jml.2011.05.002

Vasishth, S. (2003). *Working memory in sentence comprehension: Processing Hindi center embeddings*. New York: Garland Press.

Vasishth, S. (2011). Integration and prediction in head-final structures. In H. Yamashita, Y. Hirose, & J. Packard (Eds.), *Processing and producing head-final structures* (pp. 349–367). The Netherlands: Springer.

Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and anti-locality effects. *Language*, *82*, 767–794. doi:10.1353/lan.2006.0236

Wang, B., Yang, Y., & Lue, S. (2004). Hanyu yunlve cengji jiegou bianjiede shengxue fenxi [The acoustic analyses of prosodic boundaries of different levels in Chinese]. *Shengxue Xuebao* [*Acta Acustica*], *29*, 29–36. Retrieved from http://www.cqvip.com/QK/90012X/200401/9079404.html

Wightman, C. W. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *TheJournal Acoustic of Society of America*, *91*, 1707–1717. doi:10.1121/1.402450

Wu, F. 2009. *Factors affecting relative clause processing in Mandarin*. Los Angeles: University of Southern California dissertation.

Wu, F. (2011). Frequency issues of classifier configurations for processing Mandarin object-extracted relative clauses: A corpus study. *Corpus Linguistics and Linguistic Theory*, *7*, 203–227. doi:10.1515/cllt.2011.010

Wu, F., Kaiser, E., & Andersen, E. (2009). The effects of classifiers in predicting Chinese relative clauses. In the *Proceedings of the Western Conference on Linguistics (WECOL)*. Davis: University of California. Proceedings online: http://wecol.ucdavis.edu/

Wu, F., & Kaiser, E. (Resubmitted). Effects of early cues on the processing of Chinese relative clauses: Evidence for experience-based theories. *Journal of Memory and Language*.

Wu, F., Kaiser, E., & Andersen, E. (2012). Animacy effects in Chinese relative clause processing. *Language and Cognitive Processes*, *27*, 1489–1524. doi:10.1080/01690965.2011.614423

Wu, F.-Y., Haskell, T., & Andersen E. (2006). *The interaction of lexical, syntactic, and discourse factors in on-line Chinese parsing: Evidence from eye-tracking*. Poster presented at the 19th Annual CUNY Conference on Human Sentence Processing, March 23–25. CUNY Graduate Center, New York City.

Yoshida, M. (2006). Constraints and mechanisms in long-distance dependency formation (PhD thesis). University of Maryland, College Park, MD.

Yoshida, M., Aoshima, S., & Phillips, C. (2004). *Relative clause prediction in Japanese*. Paper presented at the 7th Annual CUNY conference on Human Sentence Processing. University of Maryland, College Park, MD.

Downloaded by [Peking University] at 07:40 19 October 2014