Context influences structure generation: Evidence from Chinese

Chun-chieh Natalie Hsu[1]

Felicia Hurewitz[2]

Colin Phillips[3, 4]

[1]Department of Foreign Languages and Literature, National Tsing Hua University

[2]Department of Psychology, Drexel University

[3]Department of Linguistics, University of Maryland

[4]Program in Neuroscience and Cognitive Science, University of Maryland

Address for correspondence:

Chun-chieh Natalie Hsu

Department of Foreign Languages and Literature,

National Tsing Hua University, Hsinchu, Taiwan

Email: nataliecchsu@gmail.com

Tel: +886-6-571-5131 #62027

Abstract

This study reports three sets of experiments that investigate the impact of context on structure generation in on-line sentence processing. Unlike previous studies that have addressed the role of context in resolving ambiguities, the current study examined the role of context in helping the parser to identify unambiguous structures that it could not otherwise recognize. The first set of experiments shows that late disambiguation of Chinese head-final relative clauses (RCs) elicits mild garden-path effects, and that a direct cue from the particle *suo* can facilitate the processing of head-final RCs. The second set of experiments shows that, in contrast to the direct cue, the parser fails to use an indirect-but-unambiguous cue from a mismatching classifier-noun sequence to recognize an upcoming RC structure and to avoid a garden path at the end of the RC. The third set of experiments shows that, when a supporting context for an RC is provided, the parser becomes able to use the indirect cue of the mismatching classifier-noun sequence to recognize an upcoming relative clause. Taken together, the findings suggest that there is a limitation on the parser's ability to act upon unambiguous-yet-indirect cues in incremental structure building. When such a situation occurs, contextual information becomes crucial and it actively contributes to the structure generation process to help the parser to recognize incoming structures. These findings suggest that structural and non-structural information can interact with each other in the structure generation process, providing a type of evidence for interactivity in structure generation that has proven elusive in previous studies.

Introduction

In order to understand an ambiguous sentence it is important to place the sentence in the right context. On the other hand, successful understanding of unambiguous sentences is less dependent on context. In research on on-line sentence processing, much effort has been devoted to examining how and when context can help the parser to resolve ambiguity. In contrast, there has been much less investigation of the role of context in generating unambiguous structures, possibly due to the intuition that context is not necessary in such cases. Yet this intuitive assumption is based on the observation of cases where the parser can recognize and generate unambiguous sentence structures on its own, without a supporting context. Such cases are certainly very common. Our concern in this study is with cases where the parser is unable to recognize and generate an unambiguous structure using bottom-up cues alone. In such cases we ask whether contextual support allows the parser to recognize an unambiguous structure that it could not otherwise generate. We focus on a test case from Mandarin Chinese that allows us to address exactly this question. Our findings bear on an aspect of the Modularity Hypothesis for language that has rarely been tested.

The issue of how and when context impacts the parser's operations has attracted a great deal of interest due to its relevance for Fodor's Modularity Hypothesis (Fodor, 1983). According to Fodor, a mental processing system is autonomous if it is 'informationally encapsulated', meaning that its decisions are made without counsel from higher-level processing systems or general beliefs and desires. On the other hand, a system is interactive if information from higher-level systems can guide lower-level decisions. Boland and Cutler provide an important clarification of two different ways in which higher-level information might impact language processing systems (Boland & Cutler, 1996). In the analysis of a perceptual input, they distinguish the process of *generating* candidate analyses from the process of *selecting* a preferred analysis from among those candidates. Thus contextual information might, in principle, impact the generation of candidate analyses by altering the set of analyses that are considered, or it might impact the selection among candidate analyses by changing the processor's preferred candidate, or it might have no impact on either sub-process. Boland and Cutler argue that debates about the role of context have focused on different issues in studies of word recognition and sentence parsing, despite the use of similar terminology. In lexical ambiguity resolution it is uncontroversial that context impacts

selection among word candidates, and discussions of modularity have focused on the issue of whether contextual information impacts the generation of word candidates (Becker, 1976; Borowsky & Besner, 1993; Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Morton, 1969; Norris, 1986, 1994; Pulvermüller, Assadollahi, & Elbert, 2001; Sereno, Brewer, & O'Donnell, 2003; Simpson, 1984; Smith & Besner, 2001). In contrast, debates on the role of context in sentence parsing have focused on how context impacts the selection of alternatives in ambiguity resolution (Altmann & Steedman, 1988; Chambers, Tanenhaus, & Magnuson, 2004; Crain & Steedman, 1985; Farmer, Anderson, & Spivey, 2007; Ferreira & Clifton, 1986; Frazier & Rayner, 1987; McDonald, Pearlmutter, & Seidenberg, 1994; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus & Trueswell, 1995; van Berkum, Brown, & Hagoort, 1999), or how contextual constraints can help the reader to anticipate upcoming referents (e.g. Altmann & Kamide, 1999; Kaiser & Trueswell, 2004). It is harder to find direct tests of the claim that contextual information changes the set of candidate structures that the parser generates.

In past research on the role of context in sentence processing, some studies have found weak or delayed effects of referential context (Ferreira & Clifton, 1986; Mitchell, Corley, & Garnham, 1992; Murray & Liversedge, 1994; Rayner, Garrod, & Perfetti, 1992), whereas other studies have found evidence for *early* effects of context in ambiguity resolution (Altmann, Garnham, & Henstra, 1994; Altmann & Steedman, 1988; Chambers, Tanehaus, & Magnuson, 2004; Hoeks, Vonk, & Schriefers, 2002; Liversedge, Pickering, Branigan, & Van Gompel, 1998; Ni & Crain, 1990; Spivey-Knowlton & Tanenhaus, 1994; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; van Berkum, Brown, & Hagoort, 1999). Yet another set of studies has found that referential context biases syntactic ambiguity resolution only in the absence of a strong independent bias for one of the competing syntactic alternatives (Britt, 1994; Britt, Perfetti, Garrod, & Rayner, 1992; Spivey & Tanenhaus, 1998; Spivey-Knowlton & Sedivy, 1995; Van Gompel, Pickering, & Traxler, 2000). Despite these contrasting findings, all of these studies share a focus on preferences in syntactic ambiguity resolution, and consequently they do not bear on the question of whether context can add to or delete from the set of structural candidates that the parser generates. Indeed, MacDonald and colleagues point out in an influential paper that it is unclear whether top-down

information limits the alternatives that the parser considers, since "most contexts probably are weakly constraining, in the sense that they provide an effective basis for deciding between a small number of alternatives but are less effective in isolating a single alternative in advance" (MacDonald et al., 1994, p. 697).

Advocates for the constraint satisfaction approach to parsing have pointed out the overlap between syntactic and lexical processing when resolving certain ambiguities. Many syntactic ambiguities arguably rest on lexical ambiguity. For example, the main verb/reduced relative clause ambiguity in *The man examined by the lawyer was ....* can be construed as an ambiguity between the past tense and past participle forms of the verb 'examine'. Working from this perspective, some theorists have proposed that the posited two levels of parsing should be reconsidered. According to stronger versions of this type of model, all processing occurs in the same stage, with decisions resting on the weights of different constraints, such as discourse or contextual cues, statistical likelihood of certain structures, and thematic fit between verbs and their objects. Absent from constraint satisfaction models is a clear explanation for how non-syntactic, or even non-lexical constraints can contribute to structure building or structure generation. Even these models appear to incorporate a view that in the face of unambiguous structural cues, the appropriate grammatical structure should be successfully generated on-line.

It is difficult to assess the impact of context on structure generation by investigating ambiguity resolution. This is because it is hard to empirically distinguish failure to generate a structural candidate from a mere dispreference for that candidate. An alternative strategy is to examine the processing of unambiguous structures. If a string of words has only one possible structural analysis, and if the parser's ability to recognize that single analysis is modulated by context, then this likely reflects an effect of context on structure generation. Since there is only one possible analysis, it is harder to attribute a contextual effect to the selection stage of processing. If context can alter the output of the structure generation process, then this provides a stronger argument for a non-encapsulated parser.

A number of previous studies have examined the impact of referential context on the processing of unambiguous structures. It is instructive to briefly review these studies, in order to distinguish two possible effects of contextual support on sentence processing. On one hand context may contribute to the ease or difficulty of comprehending a structure, without

changing the nature of the structural analysis that the parser constructs. On the other hand context may alter the set of structural analyses that the parser is able to generate. We suggest that the second of these scenarios would more clearly implicate a direct effect of context on the structure generation process.

A study by Grodner and colleagues showed that referential context can facilitate the processing of unambiguous sentences (Grodner, Gibson, & Watson, 2005). They found that unambiguous sentences with restrictive relative clauses (RRCs) are read more quickly when preceded by a context that supports the use of the RRC. For example, in the sentence *The postman that the dog bit on the leg needed seventeen stitches and had a permanent scar from the injury* the RRC modifier of *the postman* was easier to process when it was preceded by a context sentence that introduced a pair of postmen, i.e., *A vicious guard dog bit a postman on the leg and another postman on the arm.* Grodner and colleagues make the reasonable suggestion that the facilitation effect is due to the impact of discourse expectations. The supportive context sentence creates a potential referential ambiguity because there are two referents of the same kind in the context. The singular definite NP *the postman* in the target sentence is incompatible with the two-referent context, but this incompatibility can be resolved by the use of a RRC. Hence, reading of RRCs is facilitated by context. Grodner and colleagues interpret this effect of context on processing unambiguous structures as evidence that the discourse model directs interpretive processes and projects syntactic structures (Grodner et al. 2005, p.278). This finding is interesting, but it should be clear that context in this instance does not alter the set of structures that the parser generates - the parser can reliably recognize the RRC irrespective of context, and the effect of the supporting context is only to speed processing. This facilitation may reflect interpretive processes and be fully independent of structure generation mechanisms.

An interesting study is presented by van Berkum and colleagues (van Berkum, Brown, & Hagoort, 1999), who used ERPs to investigate the interplay of referential and structural factors during sentence processing in Dutch. They examined a case where a potentially ambiguous function word is rendered unambiguous by preceding syntactic context. In Dutch, *dat* 'that' can be a complementizer or it can be a relative pronoun associated with a neuter gender head noun. Therefore, when *dat* follows a neuter noun it is ambiguous, as it could mark the beginning of a complement clause or a relative clause. In contrast, when *dat* follows

a noun with common gender, such as *vrouw* 'woman', the relative clause parse is not possible (the common gender relative pronoun *die* would be required), so *dat* unambiguously signals the start of a complement clause. Van Berkum and colleagues measured ERPs elicited by *dat*, and manipulated the presence of two properties that make a relative pronoun construal available: the gender of a preceding noun (neuter vs. common) and contexts that introduce one or two referents described by the noun. Two referent contexts are assumed to support the use of an RRC. An example test sentence began *David vertelde {het meisje$_{NEUT}$/de vrouw$_{COM}$} dat ...* 'David told the {girl/woman} that …', and continued with either a complement clause or a relative clause. Perhaps unsurprisingly, P600 effects were elicited in the ambiguous neuter noun conditions when the sentence was disambiguated to a dispreferred structure, consistent with other reports of ERPs in garden path sentences (e.g., Brown, Hagoort, & Vonk, 1997; Kaan & Swaab, 2003; Osterhout, Holcomb, & Swinney, 1994). This was found both for a relative clause continuation in a 1-referent context and for a complement clause continuation in a 2-referent context.

Van Berkum and colleagues' most relevant finding, however, arose in the common gender conditions, where *dat* was incompatible as a relative pronoun for the preceding noun, and hence unambiguously signaled the beginning of a complement clause. If the parser relied only on bottom-up information from the incoming words, then it should have recognized the complement clause without difficulty. In fact, the ERP results showed a P600 effect at *dat* in the 2-referent condition. This suggests that speakers formed an expectation for a relative clause in the 2-referent conditions, and that this then led to disruption when the word *dat* was encountered in the input. This in turn implies that comprehenders consider analyses based on discourse context that turn out to be incompatible with the bottom-up input.

Can we conclude from this that contextual information changes the operations of the syntactic analyzer? The finding probably shows that (i) the combination of a 2-referent context and a singular definite NP can generate an expectation for a restrictive modifier, and (ii) that a mismatch is detected between this expectation and the analysis of the function word *dat*. What is less certain is whether (i) or (ii) involve actions internal to the syntactic analyzer. Property (i) could reflect an expectation for a modifier at a semantic level of analysis, with no commitment to the syntactic form of the modifier. Property (ii) could reflect a mismatch that is detected at a post-syntactic level of analysis, after the word *dat* is analyzed in a fully

bottom-up fashion as an ambiguous function word. Thus, although the study is very interesting, it does not conclusively show whether contextual information can alter the actions of the structure generator.

In light of this background, we now turn to Chinese for a case that provides a clearer test of whether discourse context can affect which representations the syntactic parser generates.

Chinese is primarily a head-initial language like English, but it has head-final noun phrases, with the consequence that Chinese has head-final RCs. Whereas English RCs follow the noun that they modify (1a), their Chinese counterparts precede the head noun (1b). The end of a Chinese RC is obligatorily marked by the relative clause marker *de*.

(1) a. the boy [$_{RC}$ who likes the cat]   *English*
    b. [$_{RC}$ xihuan na-zhi   maomi *de*] na-ge   nanhai   *Chinese*
       [$_{RC}$ like   that-CL$_{animal}$ cat   DE] that-CL   boy
         'the boy who likes that cat'

The head-final property of Chinese RCs creates potential difficulties for incremental parsing. Whereas the left-edge of English RCs is typically easy to recognize, due to overt relative pronouns (e.g., *that*, *who*) or other cues, the left edge of a Chinese RC can be much harder to identify, and hence there is greater ambiguity in Chinese than in English. For example, a noun-verb sequence like (2a) could introduce either a simple main clause (2b) or a relative clause (2c). It should come as no surprise that when speakers encounter an ambiguous noun-verb sequence like (2a) they show a general bias for a simple main clause analysis, and that a mild garden path effect arises if the sequence is disambiguated as an RC (Xiang, Munn, Schmitt, & Ferreira, 2003).

(2) a. Na-wei   xuesheng kanjian …
       that-CL$_{human}$ student   see
       'That student saw ......'
    b. Na-wei   xuesheng kanjian na-zhi   xiao-gou.
       that-CL$_{human}$ student   see   that-CL$_{animal}$ small-dog
       'That student saw that puppy.'
    c. [$_{RC}$ Na-wei   xuesheng kanjian de] na-zhi   xiao-gou   hen keai.
       that-CL$_{human}$ student   see   DE that-CL$_{animal}$ small-dog   very cute
         'The puppy which that student saw is very cute.'

Although the left edge of Chinese RCs is not normally explicitly marked, Chinese has two cues that could potentially inform the parser about the occurrence of an RC structure in advance of the RC marker *de* and the head noun. One is the particle *suo* and the other is a mismatching classifier. The particle *suo* appears optionally in Chinese RCs and is more common in the literary language, but when it does occur it unambiguously signals an object-gapped RC structure, as in (3) (Chiu, 1993, 1995; Ting, 2003). Thus *suo* provides a direct cue for a relative clause structure.

(3) [Na-wei   xuesheng   ***suo*** kanjian de]   na-zhi   xiao-gou   hen keai.
     that-CL$_{human}$ student   SUO   see   DE   that-CL$_{animal}$ small-dog   very cute
     'The puppy that the student saw is very cute.'

The other early cue for RCs in Chinese is a *mismatching classifier*. Numerals and demonstratives in Chinese must appear together with a classifier. Each classifier is associated with a certain class of nouns. For example, the classifier *ben* is used to classify book-like objects, such as dictionaries or magazines. The classifier *ben* cannot be used with other kinds of nouns (4).

(4) san-ben   gushishu   / *xuesheng
     three-CL$_{book}$ story-book /   student
      'three story-books / students'

When classifiers appear, they must occur within the noun phrase headed by their host noun. Importantly, the classifier and its host noun need not be linearly adjacent, because modifiers such as adjectives and RCs can intervene between a classifier and its host noun, as shown in (5a-b). In (5b) the head noun *jiaokeshu* 'textbook' is preceded by a modifying RC, and its classifier *ben* appears before the RC. In this case, the classifier *ben* and the adjacent noun *xuesheng* 'student' together form a mismatching classifier-noun sequence, because the classifier agrees with the head noun, but it occurs adjacent to the subject of the RC. Thus, mismatching classifier-noun sequences are grammatically possible in Chinese, but only when the sequence spans an RC boundary. Mismatching classifier-noun sequences therefore

provide an unambiguous cue for the left-edge of an RC in Chinese. However, this is a rather indirect type of cue, since neither of the lexical items in the mismatching sequence individually signal an RC structure, and nor do the two elements stand in a direct syntactic relation to one another. It is merely a quirk of Chinese grammar that mismatching classifier-noun sequences can occur only at RC boundaries.[1]

(5) a. Laoshi xihuan [$_{NP}$ na-san-*ben*   pianyi-de jiaokeshu].
    teacher like   that-three-CL$_{book}$ cheap   textbooks
    'The teacher likes those three cheap textbooks.'

    b. Laoshi xihuan [$_{NP}$ na-san-*ben*   [$_{RC}$ xuesheng feichang taoyian de] jiaokeshu].
     teacher like   that-three-CL$_{book}$ student   very   hate   DE   textbooks
     'The teacher likes those three textbooks that the students hate.'

Since a mismatching classifier-noun sequence offers a reliable-but-indirect cue to an RC boundary there is no guarantee that Chinese speakers should be able to take advantage of this cue to anticipate an upcoming RC structure and avoid a potential garden path. Although a large psycholinguistic literature supports a highly incremental sentence processing model (e.g., Frazier & Rayner 1987; Grodner & Gibson, 2005; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Lombardo & Sturt, 2002), the degree of incrementality is still under debate, especially for head-final structures (Aoshima, Phillips, & Weinberg, 2004; Aoshima, Yoshida, & Phillips, 2009; Inoue & Fodor, 1995; Kamide & Mitchell, 1999; Mazuka & Itoh, 1995; Miyamoto, 2002). In particular, the parsing actions that are needed to generate an RC structure from a mismatching classifier-noun sequence involve much more than projection of the lexical properties of the individual words. Since the classifier and the mismatching noun do not form a syntactic relation with one another, projection of their lexical properties is not sufficient to create a connected structure. More predictive parsing algorithms also do not

---

[1] This is a slight simplification, as mismatching classifier-noun sequences can arise when other types of modifiers precede a head noun, such as a noun bearing the genitive marker *-de*. However, this alternative is blocked when the mismatching noun is followed by an adverbial phrase such as *feichang* 'very', in which case only an RC structure is possible. In all of the studies reported here we took care to include adverbials, as additional cues for the RC structure.

straightforwardly succeed in this case. For example, in a standard left-corner parser (e.g., Stabler, 1994), the classifier *ben*_{*[+book]*} may license projection of an NP node that dominates the classifier and a semantically compatible noun, and the mismatching noun *xuesheng* 'student' may at best license the projection of a clausal node that dominates the subject NP *xuesheng* and a VP node. But the algorithm does not specify how to connect these two chunks. More aggressively predictive parsing algorithms may be able to successfully recognize that a mismatching classifier-noun sequence entails an RC boundary.

The specifics of the parsing algorithm do not matter for our argument here, as it suffices to observe that mismatching classifiers present unambiguous-but-indirect cues that a parser might not be able to recognize. If the parser is indeed unable to use this cue to generate a head-final RC structure, then the failure provides a test case for assessing the impact of context on structure generation. If the addition of an RC-supporting discourse context helps the parser to recognize the cue from a mismatching classifier, then this would show that context can alter the set of structures that the parser generates from a given bottom-up input. This in turn would provide a stronger case for an interactive model of structure generation (as opposed to just structure selection). We tested this logic using three sets of experiments. The first set tested whether the parser can use the direct cue from the particle *suo* to anticipate a head-final RC structure, thereby providing a baseline for the following experiments. The second set tested whether the parser can use the indirect cue from a mismatching classifier-noun sequence to recognize an upcoming head-final RC. In light of the parser's failure in those studies, the third set of studies tested whether referential context can help the parser to overcome its limitation with the indirect cue.

Experiment 1

The first experiment was designed to examine whether temporarily ambiguous head-final RCs in Chinese create difficulty in on-line sentence processing, and whether a direct cue from the particle *suo* can help the parser to avoid a garden path in head-final RCs. This study establishes a baseline and a point of comparison for subsequent experiments with indirect syntactic cues. One off-line sentence completion experiment and one on-line self-paced reading experiment were conducted.

Experiment 1a: Off-line Completion

The goal of this experiment was to check whether the particle *suo* is an effective cue for Chinese speakers to generate RCs in a setting that allows unlimited time to respond. Since *suo* is an optional element in object RCs, and is used more often in the literary language than in colloquial speech, it is possible that even a direct cue like *suo* might not be an effective cue for an RC. An additional goal of the study was to verify that Chinese speakers do not 'gratuitously' generate RCs in the absence of cues for RCs.

Methods

Participants. 40 native speakers of Mandarin Chinese spoken in Taiwan participated in this study. They were students from several different universities and institutions in Taiwan. The experiment lasted about 20 minutes.

Materials and Design. The two conditions in this sentence fragment completion study varied in the presence of the particle *suo* in the sentence fragment. Participants were shown a pencil-and-paper questionnaire consisting of sentence fragments like those in (6), and they were asked to complete each fragment in a natural manner. The dependent measure of interest was the frequency of RC structures in the fragment completions. As shown in (6), the fragments started with a subject noun phrase sandwiched between an adverbial phrase and a prepositional phrase (PP). The additional phrases provided a semantic context that made the task more natural for participants, and the PP is a good cue in Chinese for an upcoming verb. In one condition the fragments ended with the particle *suo*.

(6) Kaibanhuishi,   zhe-wei laoshi   xiang banshang-de xuesheng (suo) …
    During-class-meeting this-CL teacher   to   class-GEN   students   SUO

12 pairs of test items were distributed among two lists in a Latin Square design, such that each list contained one condition from each item set and 6 of items from each condition. Each of the two lists was combined with 24 filler items to create questionnaires with 36 items in total. Two randomizations of each list were generated to create a total of four versions of the questionnaire. The prediction was that speakers should generate few RC completions in the no-*suo* condition and a high proportion of RC completions in the *suo* condition.

Results and Discussion

The study yielded a total of 240 sentence fragment completions per condition. The completions were classified based upon the sentence structures generated by the participants. The two major classifications were 'Relative Clause' (RC) and 'others', which included simple sentences, ungrammatical sentences, and blank responses. Responses were classified as RC structures if the RC marker *de* was used. Responses were classified as others if *de* was not used. In the no-*suo* condition none of the fragments was completed with an RC structure, and in the *suo* condition 70% (167/240) of the fragments were completed with an RC structure. A Wilcoxon signed rank test showed that this difference was reliable ($N = 40$, $z = 5.16$, $p < .0001$, 2-tailed). The 30% (73/240) other responses consisted of 17.5% (42/240) ungrammatical responses and 12.5% (31/240) blank responses. 29 of the 31 blank responses were generated by only 5 participants, who gave no RC responses. Most of the ungrammatical responses were simple clause structures, and 30 of the 42 ungrammatical responses were produced by just 7 participants. These patterns indicate that the unexpected responses were produced by a relatively small subset of participants. Among the remaining 28 participants 92% (154/168) of responses in the *suo* condition included an RC structure. It is unclear why a subset of participants failed to use *suo* as a cue for RC structures. It is possible that these participants had low familiarity with *suo* because of its more literary usage. Alternatively, these participants may have paid insufficient attention to the stimulus items or may have given up when faced with difficult trials. Overall, the results of this study indicate that most Chinese speakers are able to effectively use *suo* as a direct cue for an RC structure, at least in an off-line task. We next turn to the question of whether *suo* is a similarly effective cue in an on-line setting.

Experiment 1b: Self-paced Reading

This experiment examined whether ambiguous head-final RCs create garden paths in on-line parsing in Chinese, and whether the particle *suo* can help the parser to anticipate an upcoming RC structure and hence avoid a potential garden path effect. A sample test item is shown in (7).

(7) Banhui-shi,   zhe-wei   laoshi xiang banshang-de xuesheng (suo) shenzhong-di
      class-meeting-time, this-CL teacher to   class-GEN   student   SUO   carefully

jiuzheng de   yixie guannian shi feichang zhongyaode.
correct   DE   some concept   is   very   important
'During the class meeting some concepts that the teacher carefully corrected for the students in the class are very important'

Since the RC marker *de* and the head of the RC appear at the end of the RC in Chinese, it is likely that when processing a sequence beginning with 'N V ... ' the parser initially builds a simple clause structure. Therefore, it is predicted that in the absence of other RC cues the parser should encounter difficulty upon reaching the RC marker *de* and the head of the RC, leading to reading time slowdown. If the parser is able to use the particle *suo* as a cue to an RC structure on-line, then the garden path effect should be avoided and reading times should be faster when the end of the RC is reached. Consequently, in experimental materials that manipulate the presence of *suo* in sentences like (7) we predict that reading times should be slower in the *suo* condition immediately after *suo*, since it is a low-frequency particle that cues an unexpected structure, but that the opposite pattern should be seen at the end of the RC, with faster reading times for *de* and the head of the RC in the *suo* condition. In other words, the RC cue should lead to early difficulty followed by later facilitation.

Methods

   Participants. Forty-two native speakers of Mandarin Chinese spoken in Taiwan participated in the experiment. Participants were college students studying in different institutions in Taiwan, and were naïve to the purpose of the experiment. The experiment lasted about 30 minutes.

   Materials and Design. Twenty-four pairs of items like (7) were constructed. In the *suo* condition the particle *suo* provided an early disambiguation of the RC structure, whereas in the no-*suo* condition the RC structure was not disambiguated until the appearance of the RC marker *de*. All test items included an adverbial phrase such as *xiang banshang-de xuesheng* 'to the students of the class' (cf. (7)) in order to more closely match the materials for Experiments 2-3, where the adverbial phrase was necessary. In order to increase the distance between the particle *suo* and the RC marker *de* an adverb such as *shenzhong-di* 'carefully' was added. The 24 sets of items were distributed among two lists in a Latin Square design such that each list contained 12 items from each condition. Each participant saw one of the

lists intermixed with 72 filler items in a random order. The filler items were matched with the target items in overall length and complexity. The sentences were presented in Chinese characters and the word regions were segmented in order to avoid ambiguities arising from unclear word boundaries.

Procedure. The experiment was conducted on computers running the Linger software developed at MIT by Doug Rohde. Participants were timed in a word-by-word self-paced non-cumulative moving window reading task (Just, Carpenter, & Woolley, 1982). All sentences were presented on a single line, and the segmentation indicated by spaces in the example in (7) was used in the presentation. Each trial began with a series of dashes indicating the segments and lengths of the words, and participants pressed the space bar on the keyboard to reveal each successive word. In order to ensure that participants attended to the stimuli, a yes-no comprehension question was presented at the end of each trial. Visual feedback indicating whether the answer was correct was given directly after the response. At the start of the experiment participants were given a short practice session to familiarize themselves with the task.

Data Analysis. Analyses were conducted on comprehension accuracy and reading times (RTs). All data from participants whose comprehension accuracy on the test items was less than 75% were discarded (n = 5). Reading times for all trials with incorrectly answered comprehension questions were also excluded from the analysis. Reading times that exceeded a threshold of 2.5 standard deviations above a participant's mean RT for each region were replaced by the threshold value. Three test items that had comprehension accuracy lower than 50% (21/42) were removed. In addition, two further items were removed based on ratings by 4 native speaker judges that showed the head noun to be a less natural object of the RC verb. Removal of these items did not affect the distribution of RTs but did affect the reliability of the items analysis in the critical region (R10).

Results

Comprehension Accuracy. Among the trials that were included in the analysis, the overall comprehension accuracy (including filler items) was 89.4%. The two test conditions had similar average comprehension accuracy (86.9% and 86.5%). The mean accuracy did not differ significantly across the two conditions ($t(36) = 2.03$, $p = .86$).

Reading times. Reading times at each region were entered into a repeated-measures ANOVA, with the presence of the particle *suo* (*suo* vs. no-*suo*) as the within-subjects factor. The average RT for each region in the two conditions is shown in Figure 1. Prior to the appearance of the particle *suo* in Region 7 there were no significant differences between the two conditions. At the adverb in Region 8 and the verb in Region 9 RTs were numerically longer in the *suo* condition than in the no-*suo* condition, but these differences were not statistically reliable (Region 8: *suo* = 600ms, no-*suo* = 568ms; Region 9: *suo* = 560ms, no-*suo* = 541ms). At Region 10, which included the material that disambiguated the RC (the RC marker *de*, an adjective, and the head noun), the reading time pattern was reversed, with longer RTs in the no-*suo* condition (504ms) than the *suo* condition (469ms) ($F_1(1,36) = 7.81$, $p < .01$, $F_2(1,18) = 4.57$, $p < .05$). There were no significant differences at subsequent regions.
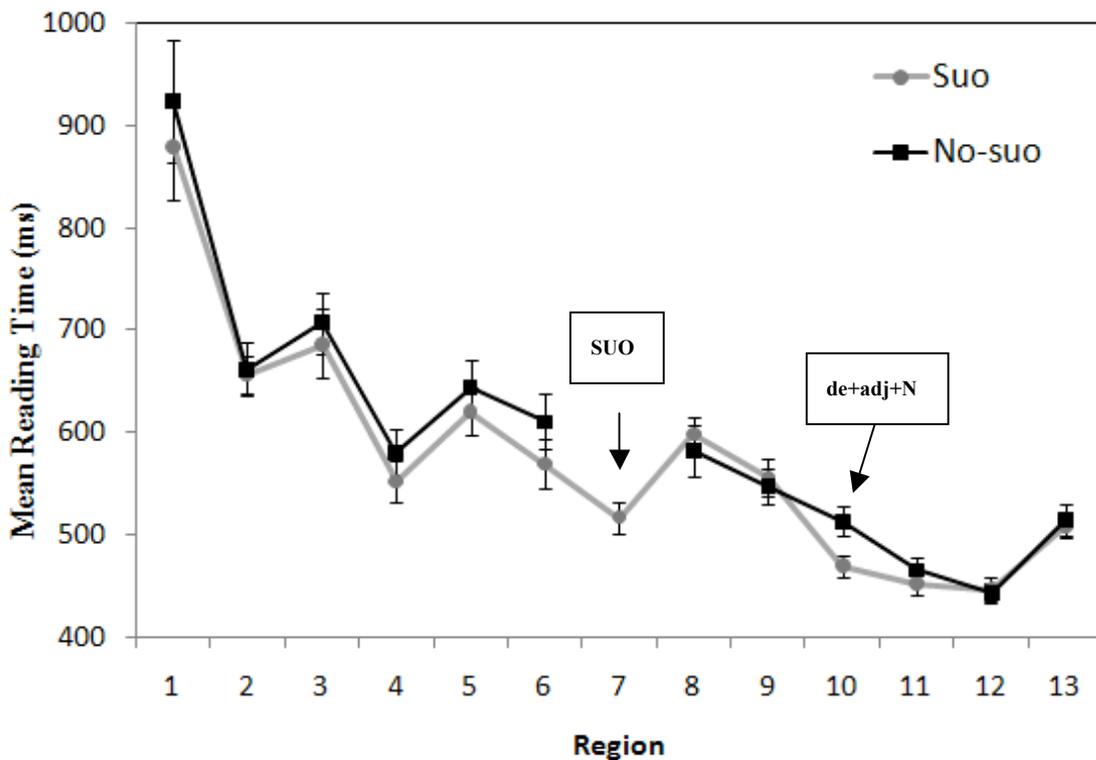
Banhui-shi$_1$, zhe-wei$_2$ laoshi$_3$ xiang$_4$ banshang-de$_5$ xuesheng$_6$ (*suo$_7$*) shenzhong-di$_8$
  class-meeting-time this-CL teacher to   class-Gen   student   SUO   carefully
    jiuzheng$_9$ de yixie guannian$_{10}$ shi$_{11}$ feichang$_{12}$ zhongyaode$_{13}$.
    correct DE some concept   is   very   important

Figure 1. Mean reading times (in ms) per region in each condition in Experiment 1b.

Discussion

The purpose of Experiment 1 was to test whether the particle *suo* can serve as an effective cue for a relative clause in both off-line and on-line tasks. Our results support the view that the presence of *suo* triggers a relative clause interpretation. In the off-line sentence fragment completion task (Experiment 1a) most participants showed a sharp contrast between the no-*suo* and the *suo* conditions, with the particle *suo* serving as an effective cue for RC completions. The results from the on-line self-paced reading study (Experiment 1b) point to a similar conclusion. Reading times showed a reversal pattern, with a non-significant trend for slower reading times in the *suo* condition immediately following the particle and then significantly faster times in that condition at the end of the relative clause. This suggests that *suo* is an effective on-line cue for an RC, and therefore helps to reduce any processing difficulty that might otherwise result from processing a head-final RC in Chinese. It should be noted, however, that ambiguous head-final RCs in Chinese do not appear to pose great difficulties for Chinese speakers, as the comprehension accuracy was high across conditions and the reading time difference at the disambiguating region was not large enough to suggest that the particle *suo* helped readers to avoid a severe garden path.

The fact that Chinese speakers are able to use the particle *suo* as an effective direct cue for a RC, leading to a reading time facilitation pattern, serves as a useful baseline for subsequent experiments that investigate Chinese speakers' ability to use indirect cues for RCs.

Experiment 2

Experiment 2 was designed to test whether the parser can use the indirect cue provided by a mismatching classifier to recognize an upcoming head-final RC. Since the classifier-noun mismatch sequence is an unambiguous-yet-indirect cue for an RC (see Introduction for discussion) it is possible that the parser might find it difficult to use this cue. As in Experiment 1 we conducted both off-line and on-line tests of the effectiveness of the cue, using different sets of participants in each case.

Experiment 2a: Off-line Fragment Completion

Methods

Participants. 36 native speakers of Mandarin Chinese spoken in Taiwan participated in this study. They were students from several different universities and institutions in Taiwan. The experiment lasted about 20-25 minutes.

Materials and Design. The two conditions in this sentence fragment completion study were identical except for the match or mismatch between a classifier and an adjacent noun. Participants were shown a pencil-and-paper questionnaire consisting of sentence fragments like those in (8), and they were asked to complete each fragment in a natural manner. The dependent measure of interest was the frequency of RC structures in the fragment completions. All target sentences began with an adverbial phrase followed by a determiner + classifier combination and a noun. The classifier varied in whether it matched or mismatched the semantic requirements of the following noun. For example, in (8) the classifier *wei* (human) is compatible with the following noun, *zuojia* 'writer', but the classifier *ben* (book) is not. A prepositional phrase following the noun provided additional semantic context and a good cue for an upcoming verb, and blocked the possibility of a possessive continuation for the classifier-noun mismatch sequence.

(8) Zai-zuotanhui-shang, {na-wei/na-ben}   zuojia yu   dangdi-de minzhong …
    During-seminar   that-CL$_{human/book}$ writer with   local   people

Twelve pairs of test items were distributed among two lists in a Latin Square design. Each list was combined with 24 filler items to create questionnaires with 36 items in total. Two randomizations of each list were generated to create a total of four versions of the questionnaire. The prediction was that speakers should generate few RC completions in the classifier-match condition, and that they should generate many more RC completions in the classifier-mismatch condition if the mismatch is an effective indirect cue for a relative clause.

Results and Discussion

The study yielded a total of 216 sentence fragment completions per condition. We classified the completions based upon the sentence structures produced by the participants. The two major classifications were 'Relative Clause' (RC) and 'others,' which included

simple sentences, ungrammatical sentences, and unfinished or blank responses. Responses were classified as RC structure if the RC marker -de was used.

In the classifier-match condition none of the fragments was completed with an RC structure, and in the classifier-mismatch condition 44% (94/216) of the fragments were completed with an RC structure. A Wilcoxon signed rank test showed that this difference was reliable ($N = 36$, $z = 4.45$, $p < .0001$, 2-tailed). Nevertheless, there was a high proportion of 'other' responses in the classifier-mismatch condition. These responses consisted of 50% (108/216) ungrammatical responses (primarily simple clause continuations) and 6% (14/216) NP-deletion responses. Representative examples are shown in (9).

(9) a. Ungrammatical responses: simple clause continuation

*Zai-zuotanhui-shang, na-ben   zuojia yu   dangdi-de minzhong yichi   hudong
During-seminar   *that-CL_{book} writer with local   people   together interact*
'During the seminar, the writer and the local people interacted together.'

b. NP-deletion response

Zai-zuotanhui-shang, na-ben,   zuojia yu   dangdi-de minzhong
During-seminar,   *that-CL_{book}   writer with   local   people*
zhuo-le yixie   taoluan.
make   some discussion
'During the seminar, the writer had some discussion with the local people about that one (book).'

The NP-deletion continuation requires a pragmatic context that allows topicalization (Chen, 1996). An object NP can be topicalized when it is emphasized, and if an appropriate context is provided, the topicalized noun can be deleted. In the example in (10), the noun *shu* 'book' is introduced in the question and it undergoes topicalization and deletion in the response. Deletion structures were rare in the sentence completion data, and 8 of the 14 responses of this type were contributed by 3 participants.

(10) a. Shei feichang xihuan na-san-ben   shu?
who   very   like   that-3-CL_{book} book
'Who likes those three books?'

b. Na-san-*ben*,　　　*xuesheng* feichang xihuan.

　　that-3-CL*book*　　*student*　　*very*　　*like*

　　'For those three (books), the student likes (them) very much.'

The results from the sentence fragment completion task indicate that a mismatching classifier-noun sequence, which provides an unambiguous-yet-indirect cue to an RC structure, can induce Chinese speakers to produce RC structures to some degree in an off-line setting. However, the high number of ungrammatical completions (around 50%) suggests that the cue is not uniformly salient for Chinese speakers. As in the previous sentence completion task (Experiment 1a) the ungrammatical responses were not evenly distributed across participants, with 66/108 ungrammatical responses contributed by 12 participants. This means that 12 participants produced ungrammatical responses in over 90% of trials, and the remaining 24 participants produced fewer than 30% ungrammatical responses. There are multiple possible explanations of this uneven distribution of RC responses. It is possible that the classifier mismatch cue is simply a difficult cue to use, and that it lies beyond the parsing capacity of some Chinese speakers to use such an indirect cue. Alternatively, it is possible that some participants did not attend sufficiently to the details of the sentence fragments (among which the classifier-mismatch items made up just one sixth of trials), or assumed that the questionnaire contained errors. Due to the nature of the task, participants never received feedback showing that an RC completion was a grammatical completion of the classifier-mismatch fragments. Nevertheless, these data do show that for most Chinese speakers the classifier-mismatch cue is at least somewhat effective in an off-line setting. We next investigate whether it is as effective as the direct RC cue *suo* in an on-line setting.

Experiment 2b: On-line Reading

Experiment 2a showed that mismatching classifier-noun sequences serve as moderately effective cues for RC structures in an off-line setting. The goal of the current study was to test whether this unambiguous-yet-indirect cue supports early recognition of Chinese RC structures in an on-line setting, in the same way that the direct cue *suo* did in Experiment 1b. All parsing models that we are aware of predict that upon encountering a mismatching classifier-noun sequence Chinese speakers should initially encounter difficulty, as such sequences are rare, unexpected, and are associated with complex structures. However, if

speakers are able to take advantage of this cue to recognize an RC structure, then we should find a reversal in reading time patterns at the end of the RC (the particle *de* and the head noun), since this evidence for an RC structure would be expected in the classifier-mismatch condition but unexpected in the classifier-match condition. Alternatively, if speakers cannot use the indirect cue to an RC we should still find the initial slowdown at the mismatching classifier-noun sequence, but no reading time reversal at the end of the RC.

Method

   Participants. 16 native speakers of Mandarin Chinese spoken in Taiwan participated in the experiment. None of them were involved in the previous experiments and they were naïve to the purpose of the study. All were students at the University of Maryland. The experiment lasted around 30 minutes.

   Materials and Design. A sample set of test conditions is shown in (11). All test items contained RCs, but they varied in the presence of an early cue to the RC in the form of the mismatching classifier-noun sequence. In the classifier-mismatch condition (11a) the classifier (e.g., *pian* 'article') was incompatible with the adjacent noun (e.g., *xuesheng* 'student'), and was only compatible with the later head noun (e.g., *wenzhang* 'article'). In the classifier-match condition (11b) the classifier was instead congruent with the adjacent noun (e.g., *wei* 'human') and therefore it did not provide a cue for a RC structure. The RC marker *de* and the head noun always marked the end of the RC. In anticipation of a long-lasting disruption caused by the classifier-noun mismatch, which might obscure effects of interest at the end of the RC, four regions were added between the subject of the RC (*xuesheng* 'student' in (11)) and the embedded verb position (*taolun-de* 'discuss-DE' in (11)). These intervening regions contained unambiguous adverbial phrases, such as frequency adverbs, manner, comitative, or purpose adverbial phrases, none of which could function as predicates by themselves, and none of which contributed additional RC cues.

(11) a. *Classifier-mismatch Condition*
       Laoshi caixiang na-si-*pian*   yonggong-de xuesheng zaoshang jingchang yu
       teacher guess-think na-four-CL$_{article}$ *diligent student   morning   often   with*
       tongxue   taolun-de   wenzhang yinggai   dui   xiezuo   youbangzhu.
       classmate discuss-DE   article   should   to   writing   helpful
       'The teacher thinks that the four articles that the diligent student often discusses with a
        classmate in the morning should be helpful to their writing.'

b. *Classifier-match Condition*
  Laoshi caixiang   na-si-*wei*   yonggong-de xuesheng zaoshang jingchang yu
  teacher guess-think na-four-CL$_{human}$ diligent   student   morning   often   with
  tongxue   taolun-de   wenzhang yinggai dui   xiezuo   youbangzhu.
  classmate discuss-DE   article   should   to   writing   helpful
  'The teacher thinks that the articles that the four diligent students often discuss with a
   classmate in the morning should be helpful to their writing.'

Twenty-four pairs of two conditions were distributed among two lists in a Latin-Square design. Each participant saw exactly one of the lists intermixed with seventy-two filler items of similar length and complexity in a pseudorandom order. The procedure was the same as for Experiment 1b.

  Data Analysis. Analyses were conducted on comprehension accuracy and reading times (RTs). Since all participants scored above 75% on the comprehension task in this study, no participants were removed from the analysis. All trials for which the comprehension question was answered incorrectly were also excluded. Reading times that exceeded a threshold of 2.5 standard deviations above a participant's mean RT for each region were replaced with the threshold value. The means and analyses presented below are based on the remaining trials.

Results

  Comprehension Accuracy. The overall comprehension accuracy on all trials was 91%. Within the target items comprehension accuracy was higher in the classifier-mismatch condition (87.5%) than in the classifier-match condition (81.8%), although the difference was only marginally significant ($t(15) = 2.13$, $p = .068$). Accuracy for the filler items was 93.3%.

  Reading times. Reading times at each region were entered into a repeated-measures ANOVA, with semantic compatibility between the classifier and the embedded subject (mismatch vs. match) as a within-subjects factor. Average reading times for each region are shown in Figure 2. There were no differences in Regions 1-3. At the embedded subject noun phrase in Region 4 (the adjective) and Region 5 (the embedded noun), RTs were longer in the classifier-mismatch condition, but the difference did not reach significance (Region 4: classifier-match = 571ms, classifier-mismatch = 615ms; Region 5: classifier-match = 702ms, classifier-mismatch = 767ms). At Regions 6-8, which contained the adverbial phrases that followed   the   classifier-noun   mismatch,   RTs   were   significantly   longer   in   the

classifier-mismatch condition: Region 6 (classifier-match = 537ms, classifier-mismatch = 746ms; $F_1(1,15) = 11.38$, $p < .01$; $F_2(1,23) = 17.21$, $p < .001$); Region 7 (classifier-match = 480ms, classifier-mismatch = 577ms; $F_1(1,15) = 17.68$, $p < .001$; $F_2(1,23) = 14.52$, $p < .001$); Region 8 (classifier-match = 502ms, classifier-mismatch = 598ms; $F_1(1,15) = 7.56$, $p < .05$; $F_2(1,23) = 6.62$, $p < .05$). There was no slowdown at the last word of the adverbial phrase in Region 9, which appeared directly before the critical regions. At the critical regions at the end of the RC there was no reversal in reading time patterns. At the embedded verb and RC marker *de* in Region 10 there was no reliable difference between conditions (classifier-match = 481ms, classifier-mismatch = 530ms), and at the head noun in Region 11 RTs were again slower in the classifier-mismatch condition (classifier-match = 521ms, classifier-mismatch = 601ms; $F_1(1,15) = 5.37$, $p < .05$; $F_2(1,23) = 5.63$, $p < .05$).
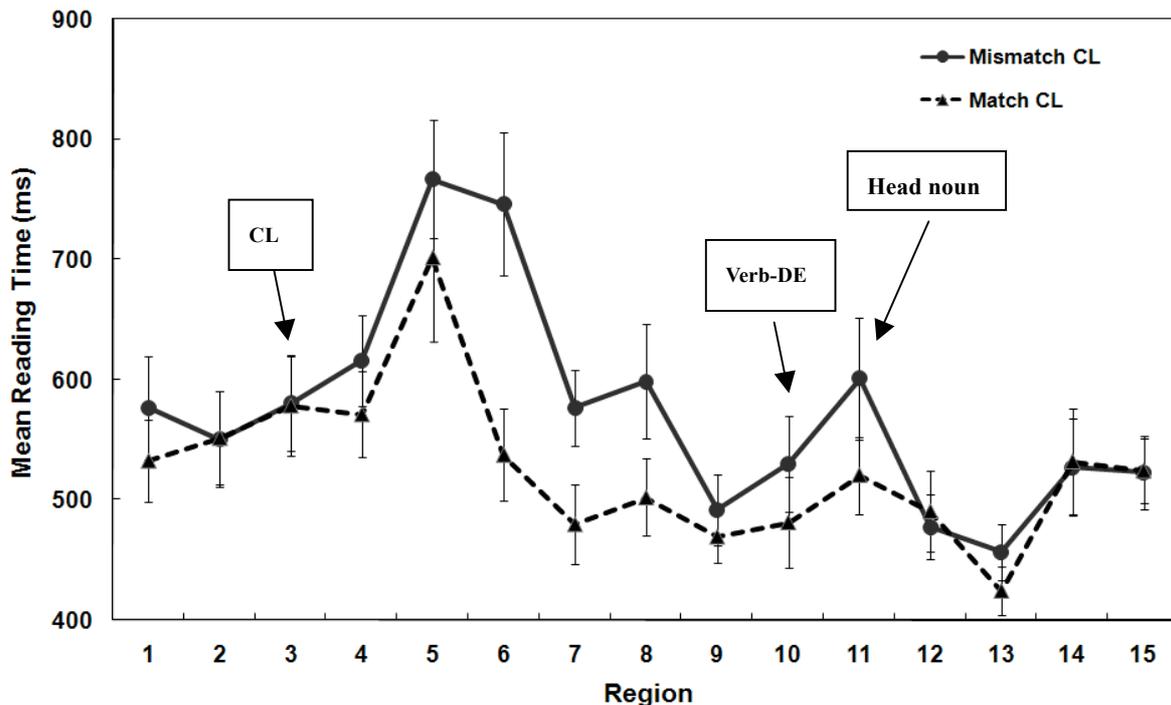
Laoshi₁ caixiang₂   na-si- *wei₃*/*pian₃*   yonggong-de₄ xuesheng₅ zaoshang₆ jingchang₇   yu₈
teacher guess-think na-four-CL_{human/article}   diligent   student   morning   often   with
tongxue₉   taolun-de₁₀   wenzhang₁₁ yinggai₁₂ dui₁₃   xiezuo₁₄   youbangzhu₁₅.
classmate   discuss-DE   article   should   to   writing   helpful

'The teacher thinks that the four articles that the diligent student often discussed with a classmate in the morning should be helpful to their writing.'

Figure 2. Mean reading times (in ms) per region in Experiment 2b

Discussion

The results of this experiment tell us several important things. Firstly, the slowdown at the embedded subject position that lasted for several regions (Regions 5-8) in the classifier-mismatch condition confirms that the classifier-noun mismatch sequence led to substantial processing difficulty. This slowdown is not surprising, but it could not be taken for granted, particularly in light of the relatively mild difficulty encountered by Chinese speakers in processing head-final RCs in general (see Experiment 1b). The early slowdown will be referred to as the 'mismatch effect' in what follows. Secondly, at the critical regions at the end of the RC (Regions 10-11) there was no reversal in the pattern of reading difficulty. Instead, RTs at the head noun position were higher in the classifier-mismatch condition than in the classifier-match condition. This lack of reversal in reading time patterns contrasts with the pattern observed following manipulation of the RC cue *suo* in Experiment 1b, and it suggests that Chinese speakers were not able to use the classifier-mismatch cue as an early recognition cue for RC structures. We speculate that the elevated reading times at the head noun position in the classifier-mismatch condition may reflect the cost of long-distance integration between the mismatching classifier and the head noun (cf. Gibson, 1998, 2000).

We conducted three other similar experiments using different sets of adverbial phrases and verbs, and the results of these experiments were consistent with one another. All of the experiments elicited a mismatch effect at the mismatching classifier-noun sequence, and none of the experiments elicited a reversal in reading time patterns at the end of the RC. (There was inconsistency in whether studies showed no differences between conditions at the end of the RC or a mismatch cost as seen in the current study.) We do not report these other studies in detail here as they all lead to the same conclusion about the effectiveness of the classifier mismatch cue in an on-line setting. It appears that there is a limit on the parser's ability to take advantage of this indirect cue. There are a number of possible explanations for this limitation, involving such factors as the low frequency of the mismatching sequences, the limited structure-projection capacities of left-corner and related parsers (Abney & Johnson, 1991; Johnson-Laird, 1983; Stabler, 1994; Sturt & Crocker, 1996), and the markedness of RC structures in general. Further work will be needed to pinpoint the source of the limitation, and it will also be important to identify why very similar RC cues appear to be more effective in Japanese, a more consistently head-final language than Chinese (Yoshida, Aoshima, &

Phillips, 2004). However, for purposes of the current argument it is sufficient to have identified that in mismatching classifier-noun sequences, the parser is not able to generate and construct an unambiguous structure based on bottom-up information alone. This structural configuration therefore offers a promising candidate for investigating the role of context in structure generation. As discussed in the Introduction, it is difficult to draw firm conclusions about the role of context in structure generation from cases where the parser is able to recognize a structure irrespective of context, or from cases of structural ambiguity, since effects of context in these situations may reflect mechanisms other than structure generation (e.g., structure selection, discourse integration). However, the case of classifier-noun mismatches are a good test case: they provide an unambiguous cue to an RC structure, but Experiment 2b indicates that Chinese speakers are not able to use this cue effectively to facilitate on-line processing of head-final RCs in the absence of a supportive context. This sets the stage for investigating whether the addition of a supportive context allows Chinese speakers to use the classifier-mismatch cue on-line. If context does render the cue effective, then this implies that it specifically impacts the Chinese parser's structure generation actions. Experiment 3 addresses this question.

Experiment 3

This experiment examined whether the addition of a supportive context for an RC can assist the parser to use the unambiguous-yet-indirect cue from a mismatching classifier to generate an RC structure. If the parser is able to use the mismatch cue to produce an RC structure only in an RC-supporting context, but not in a non-RC-supporting context, then the result would suggest that context directly impacts the parser's structure generation mechanism.

Relative clauses can be used in a wide variety of discourse contexts, but one function has received particular attention in the sentence processing literature. Restrictive relative clauses are used to differentiate referents of the same type. Therefore use of a restrictive relative clause in a single referent context is thought to violate the pragmatic expectations of discourse, where redundant information is only used to impart some additional meaning (Grice, 1957). In accordance with Grice's Maxim of Quantity, listeners only expect restrictive relative clauses in situations where a simple NP would not uniquely refer to single referent, as

otherwise the RC is informationally redundant. Numerous findings on the processing of structurally ambiguous sentences have shown that when restrictive relatives are used in single referent or "null" contexts, this pragmatic infelicity creates processing difficulties for the comprehender. Studies in this tradition have further investigated how RC-supporting contexts can guide structural ambiguity resolution, to allow the processing of infrequent or more complex structures without a garden path effect as reflected in responses or reading times (Altmann & Steedman, 1988; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; van Berkum et al., 1999), or to facilitate the processing of easy-to-recognize unambiguous structures (Grodner et al., 2005).

In this next set of studies we manipulated the presence of an RC-supporting context and the presence of a classifier mismatch cue for an RC. In RC-supporting contexts the discourse context provided two referents of the same type that were compatible with the classifier, and the aim was to test whether the supportive context would render the classifier mismatch cue effective. We conducted two off-line sentence fragment completion experiments and two on-line self-paced reading experiments on this issue.

Experiment 3a: Off-line Completion

The goal of this experiment was to provide a baseline test of whether an RC-supporting context induces Chinese speakers to generate RCs in the absence of syntactic or lexical cues such as the particle *suo* or mismatching classifier-noun sequences. In this sentence fragment completion study the last region of the target fragments was a singular determiner + classifier combination, and participants were not shown classifier + noun sequences. This made it possible to test whether the combination of a multiple referent context and a singular determiner would bias speakers to generate RC completions in the absence of syntactic cues.

Methods

Participants. 31 native speakers of Mandarin Chinese spoken in Taiwan participated in this study. They were students from several different universities and institutions in Taiwan. The experiment lasted about 20 minutes.

Materials and Design. The two conditions in this sentence fragment completion study contained identical fragments, but differed in the content of a context sentence that preceded

each target sentence. Participants were shown a paper-and-pencil questionnaire consisting of context sentences and fragments like those in (12), and they were asked to complete each fragment in a natural manner. As in the previous fragment completion studies in this article the dependent measure of interest was the frequency of RC structures in the fragment completions. The context sentences in the 2-referent and 1-referent conditions were very similar, as shown in (12). The 2-referent condition (12a) introduced two referents of the same kind, which matched with the classifier in the target fragment (12c), whereas the 1-referent condition (12b) introduced one referent that matched the classifier in the target fragment and another that did not. The target fragments were shorter than in other studies in this article, containing just an introductory adverbial phrase and a determiner + classifier combination.

(12) a. *2-Referent Condition*

    Loushangde daxuesheng   you *liang-liang   motuoche*. Ta wanquan meiyou
    upstairs   college-student   have two-CL$_{vehicle}$ motorcycle he totally   not
    baoyang   qizhongde   yi-liang   motuoche, danshi ta que hen yongxindi
    maintain   among   one-CL$_{vehicle}$ motorcycle but   he yet very carefully
    baoyang   lingwai yi-liang   motuoche.
    maintain   another   one-CL$_{vehicle}$ motorcycle
    'The college-student upstairs has two motorcycles. He does not maintain one of the motorcycles at all, but he maintains the other motorcycle very carefully.'

   b. *1-Referent Condition*

    Loushangde daxuesheng   you *yi-liang   motuoche han yi-tai   diannao.*
    upstairs college-student have one-CL$_{vehicle}$ motorcycle and one-CL$_{device}$ computer
    Ta meiyou yongxing baoyang tade diannao, danshi ta que hen yongxindi baoyang
    He   not   carefully maintain his computer but   he yet very carefully maintain
    Tade motuoche.
    his motorcycle
    'The college-student upstairs has one motorcycle and one computer. He does not carefully maintain his computer, but he maintains his motorcycle very carefully.'

   c. Zhege   xueqi-zhong,   na-liang ...
    this   semester-during that-CL$_{vehicle}$ …
    'During this semester, that …'

Twenty pairs of test items were distributed among two lists in a Latin Square design, such that each list contained one condition from each item set and 10 items from each condition. Two randomizations of each list were generated to create a total of four versions of the questionnaire. Each questionnaire was then combined with 40 fillers to create lists of 60 items in total. The lists were split in two such that each participant completed only half of one list. The prediction was that there should be more RC completions in the 2-referent context than in the 1-referent context if Chinese speakers are able to recognize that an RC modifier makes it possible to use a singular determiner + classifier combination in a context that contains multiple referents of the same type that match the classifier.

Results and Discussion

The study yielded a total of 155 sentence fragment completions per condition. We classified the completions based upon the sentence structures generated by the participants. Responses were classified into the categories 'Relative Clause' (RC) and 'others', using the same criteria as in Experiments 1a and 2a. In the 2-referent condition 83% (128/155) of the fragments were completed with an RC structure whereas in the 1-referent condition only 5% (7/155) of the fragments were completed with an RC structure. A Wilcoxon signed rank test confirmed that this difference was reliable ($N = 40$, $z = 4.86$, $p < .0001$, 2-tailed). In the 2-referent condition the 17% of responses in the 'other' category consisted of 14% (22/155) simple clause responses (the 22 responses were contributed by 16 participants), 1% (2/155) unfinished/blank responses, and 2% (3/155) noun deletion responses. In the 1-referent condition the 95% of responses in the 'other' category included 94% (146/155) simple clause structures.

These results indicate that Chinese speakers generate significantly more RC completions for fragments containing a singular determiner + classifier combination when that combination appears in a context containing multiple referents of the same kind that match the classifier. This suggests that, at least in an off-line setting, just the combination of a multiple referent context and an associated singular determiner is sufficient to favor generation of RC structures. This finding is relevant to the overall line of argumentation in this article, since the sentence fragments in this study contained no syntactic cues at all for RC structures, either direct or indirect. If RC structures can be generated using joint cues

from two semantic properties (2-referent context, singular determiner), then care must be taken in constructing tests of interactivity and modularity in structure generation. Only if it can be shown that the generation of a structure crucially requires the combination of syntactic and semantic cues is it possible to construct an argument for interactivity in structure generation.

Experiment 3b: Off-line Completion Experiment

The goal of the second sentence fragment completion study in this set was to provide a first test of the impact of context on the effectiveness of the indirect syntactic cue provided by classifier-noun mismatches. In contrast to Experiment 2a this study did not manipulate the match between the classifier and the following noun, using only target fragments with mismatching classifier-noun sequences.

Methods

Participants. Thirty-eight native speakers of Mandarin Chinese spoken in Taiwan participated in this study. They were students from several different universities and institutions in Taiwan. The experiment lasted about 20 minutes.

Materials and Design. The materials and design of this study were identical to Experiment 3a, except that a noun was added to the target fragments to create a mismatching classifier-noun sequence, as illustrated in (13). As in Experiment 3a the two conditions varied only in nature of the context sentence, which provided either one or two referents of the same kind that were compatible with the classifier in the target fragment.

(13) Zhege xueqi-zhong,   na-*liang*   daxuesheng …
    this   semester-during   that-CL$_{vehicle}$ college-student …

Twenty sets of test items were distributed among two lists in a Latin Square design, such that each list contained one condition from each item pair and 10 items from each condition. Two randomizations of each list were generated to create a total of four versions of the questionnaire. Each questionnaire was then combined with 40 fillers to create lists of 60 items in total. The filler items included a number of matching classifier-noun sequences, but these appeared only in the context sentences and never in the sentence fragments. The lists

were split in two such that each participant completed only half of one list. The prediction was that there should be more RC completions in the 2-referent context than in the 1-referent context if Chinese speakers are able to use the combined cues from an RC-supporting context and a classifier-noun mismatch to generate RC structures. However, since we have already observed that these two cues individually facilitate RC generation in an off-line setting (classifier-noun mismatch - Experiment 2a; 2-referent context - Experiment 3a) it is possible that the cue combination may provide a limited additional benefit.

Results and Discussion

This study yielded a total of 190 sentence fragment completions per condition. The completions were classified into the categories 'Relative Clause' (RC) and 'others', using the same criteria as in the previous completion studies. In the 2-referent condition 76% (145/190) of the fragments were completed with a RC structure and in the 1-referent condition 69% (131/190) were completed with a RC structure. A Wilcoxon signed rank test showed that this difference was not reliable ($N = 40$, $z = 1.49$, $p = 0.14$, 2-tailed). In the 2-referent condition the 24% of responses in the 'other' category consisted of 13% (24/190) ungrammatical simple clause continuations (contributed by 15 participants; among the 24 responses, 14 were from the same 5 participants.), 8% (16/190) NP-deletion constructions (contributed by 6 participants; among the 16 responses, 13 were from the same 3 participants.), 2% (3/190) possessive constructions, and 1% (2/190) blank responses. In the 1-referent condition the 31% of responses in the 'other' category consisted of 13% (25/190) ungrammatical simple clause continuations (contributed by 15 participants; among the 25 responses, 18 of them were from the same 8 participants.), 5% (9/190) NP-deletion constructions (contributed by 4 participants; among the 9 responses, 7 were from the same 2 participants.), 12% (23/190) possessive constructions (contributed by 18 participants; among the 23 responses, 10 were from the same 4 participants.), and 1% (2/190) blank responses. Further statistical tests showed that more possessive continuations were generated in the 1-referent condition than in the 2-referent condition. ($N = 38$, $z = 3.625$, $p < 0.001$, 2-tailed).

This experiment showed that in an off-line setting the manipulation of the semantic context preceding a classifier-noun mismatch did not have a significant effect on the proportion of generated RC structures. This is perhaps not surprising, since Experiment 2a

already showed that a classifier-noun mismatch is an effective cue for RC structures even in the absence of an RC-supporting context. Nevertheless, since Experiment 2b showed that the effectiveness of the classifier-mismatch cue did not extend to an on-line setting, it remains very relevant to ask whether the combined cues are effective in an on-line task (Experiment 3c).

Comparison of the rates of RC structures in the different sentence completion studies reported here indicates some variability for similar conditions. The overall rate of RC completions for classifier-noun mismatches was 76% in Experiment 3b and only 44% in Experiment 2a. The 76% rate of RC completions in 2-referent contexts in Experiment 3b was lower than the 83% rate in the corresponding condition in Experiment 3a, although this difference was not reliable. These differences may reflect the details of the participants or the task used in the different studies. Recall that the rate of RC completions in Experiment 2a was closer to 70% among the majority of participants who recognized the classifier-mismatch cue, since most errors in that study were produced by a minority of participants. Overall rates of RC completion may also have been elevated in Experiment 3a due to the absence of a classifier-match condition in this study. In this off-line experiment, the greater proportion of trials that called for an RC completion may have increased the salience of the classifier-mismatch cue and may have led to greater priming of RC structures.

Although the difference between the number of RC completions in the 1-referent and 2-referent conditions in the current study was not significant, the slightly lower rate of RC completions in the 1-referent condition likely reflects the higher number of possessive constructions generated in that condition. As discussed above, the possessive construction is a possible continuation of a classifier-noun mismatch sequence in the absence of additional adverbial phrases. This continuation involves simply adding the possessive morpheme -de to the classifier-noun mismatch sequence together with a noun that matches the classifier, as illustrated in (14). The possessive structure is likely easier than an RC structure. The finding that Chinese speakers generated more possessive continuations in the 1-referent context (12%) than in the 2-referent context (2%) may reflect the fact that the 1-referent context made it slightly more difficult to recognize the indirect RC cue, with the consequence that some participants generated possessive structures instead. Note that the addition of an adverbial phrase following the mismatching classifier-noun sequence blocks the possessive

construction. An adverbial phrase of this kind was included in offline Experiment 2a and in all of the on-line studies of classifier-noun mismatch processing. It was not included in Experiment 3b, in order to minimize the differences between Experiments 3a and 3b.

(14) Zhege xueqi-zhong,   na-*liang*   daxuesheng-*de*   motuoche bei-tou-le.
     this semester-during that-CL$_{vehicle}$ college-student-POSS motorcycle BEI-steal-ASP.^{2}
     'During this semester, that college student's motorcycle was stolen.'

The current results are also informative regarding the salience of the NP-deletion construction. As discussed earlier, the NP-deletion construction is possible, but it requires a special pragmatic context that licenses a combination of topicalization and an available antecedent for the deleted noun. The proportion of NP-deletion continuation responses was relatively small in this study, at 6.6% (25/380) overall. Furthermore, these few responses were confined to a small number of participants: 13 of the 16 NP-deletion responses were contributed by 3 participants in the 2-referent condition, and one of these participants accounted for 5 of the 9 NP-deletion responses in the 1-referent condition. We can therefore conclude that for the vast majority of Chinese speakers the NP-deletion construction is not a salient analysis of mismatching classifier-noun sequences.

Having confirmed the effect of RC-supporting contexts in off-line tasks, we next turn to examine whether RC-supporting contexts can overcome the ineffectiveness of the classifier-mismatch cue in an on-line setting.

Experiment 3c: On-line Reading

This experiment was designed to examine whether a felicitous context for RCs can help the parser to exploit the indirect cue from a mismatching classifier to recognize an RC structure in an on-line setting. In this self-paced reading experiment all items consisted of a context sentence followed by a target sentence that contained a head-final relative clause. The experiment manipulated the number of referents in the context sentence that matched the classifier in the target sentence, and the match between the classifier and an adjacent noun in the target sentence. We have already seen in Experiment 2b that Chinese speakers have difficulty using the classifier mismatch cue to a RC online in the absence of a supportive

---

^{2} POSS = Possessive; ASP = Aspect marker

context. If a supportive discourse context makes the classifier mismatch cue more effective, then we expect to see the same cross-over pattern in reading times observed in Experiment 1b for the direct RC cue from the particle *suo*: reading times should be slower immediately after the mismatching classifier, but should show a reversal with faster reading times at the head of the RC.

Methods

Participants. Twenty-eight native speakers of Mandarin Chinese spoken in Taiwan participated in the experiment. They were graduate students at the University of Delaware and were naïve to the purpose of the experiment. The experiment lasted about 30 minutes.

Materials and Design. The experiment had a 2 × 2 factorial design that manipulated the factors context (1-referent vs. 2-referent) and classifier-match (classifier-match vs. classifier-mismatch). Most of the text materials were very similar to those used in the off-line studies (Experiments 3a-b). The context sentences were the same as in the sentence fragment completion studies (12a-b above, repeated below in 15a-b). The target sentences all contained a head-final relative clause, and this was signaled by a mismatching classifier-noun sequence in the classifier-mismatch condition (16a), a cue that was lacking in the classifier-match condition (16b). Note that the relation between the classifier and the context manipulation was different in the classifier-match and the classifier-mismatch conditions. In the classifier-match conditions the classifier (e.g., *wei* 'human' in (16b)) matched the following noun, which was the single agent in both context conditions. In the classifier-mismatch conditions the classifier (e.g., *liang* 'vehicle' in (16a)) matched a single patient referent from the context sentence in the 1-referent condition and two patient referents in the 2-referent condition.

(15) a. *2-Referent Condition*
Loushangde daxuesheng   you *liang-liang   motuoche*. Ta wanquan meiyou
upstairs   college-student   have two-CL$_{vehicle}$ motorcycle he totally   not
baoyang   qizhongde   yi-liang   motuoche, danshi ta que hen yongxindi
maintain   among   one-CL$_{vehicle}$ motorcycle but   he yet very carefully
baoyang   lingwai   yi-liang   motuoche.
maintain   another   one-CL$_{vehicle}$ motorcycle
'The college-student upstairs has two motorcycles. He does not maintain one of the
 motorcycles at all, but he maintains the other motorcycle very carefully.'

b. *1-Referent Condition*

Loushangde daxuesheng you *yi-liang   motuoche   han yi-tai   diannao.*
upstairs college-student have one-CL$_{vehicle}$ motorcycle and one-CL$_{device}$ computer
Ta meiyou yongxing baoyang tade diannao, danshi ta que hen yongxindi baoyang
He not   carefully maintain his computer but   he yet very carefully maintain
tade motuoche.
his motorcycle
'The college-student upstairs has one motorcycle and one computer. He does not
carefully maintain his computer, but he maintains his motorcycle very carefully.'

(16) a. *Classifier-mismatch Condition*

Zhege xueqin-zhong,   na-**liang**   daxuesheng   chang-yi
this   semester-during   that-CL$_{vehicle}$ college-student   often-with
xixinde taidu baoyang-de motuoche jingran bei-tou-le.
careful   attitude maintain-DE   motorcycle suddenly was-stolen-ASP

b. *Classifier-match Condition*

Zhege xueqin-zhong,   na-**wei**   daxuesheng   chang-yi
this   semester-during that-CL$_{human}$ college-student often-with
xixinde taidu   baoyang-de   motuoche   jingran   bei-tou-le.
careful attitude maintain-DE motorcycle suddenly was-stolen-ASP
'During this semester, the motorcycle which the college student often carefully
maintains was suddenly stolen.'

As in Experiment 2b the classifier-noun sequence in the target sentences was followed by a series of adverbial phrases that served multiple purposes. The adverbials served to block the possessive construction parse of the mismatching classifier-noun sequences, and they also could not be analyzed as predicates by themselves and required a following verb. In addition, the adverbials served to reduce the possibility that continuing reading time disruption caused by the classifier-noun mismatch sequence would overlap with effects of interest at the disambiguating verb and head noun regions at the end of the RC. All of the RCs in this study were modifiers of the main clause subject, and this avoided any possible additional ambiguities caused by the main clause verb. 20 sets of four conditions were distributed among 4 lists in a Latin Square design. Each participant saw one of the lists intermixed randomly with 64 filler items of similar length and complexity. The procedure for this experiment was identical to that used in the self-paced reading experiments described above.

Data Analysis. Analyses were conducted on comprehension accuracy and reading times. All participants in this study had comprehension accuracy above 75%, so no participants were excluded from the analysis. Reading times for the sentences whose comprehension answers were incorrect were also excluded. Reading times that exceeded a threshold of 2.5

standard deviations above a participant's mean RT for each region were replaced by the threshold value. The means and analyses presented below are based on the remaining trials.

Results

Comprehension Accuracy. The overall comprehension accuracy was 95.5%, and the average comprehension accuracy for target items was 97.7%. The average comprehension accuracy was 95.0% in the 2-referent/classifier-match condition and 98.7% in the other three conditions. The differences were not reliable.

Reading times. Reading times at each region were entered into a repeated-measures ANOVA, with the factors context (2-referent vs. 1-referent) and the classifier (classifier-match vs. classifier-mismatch) as within-subjects factors. Average reading times for each region are shown in Figure 3.
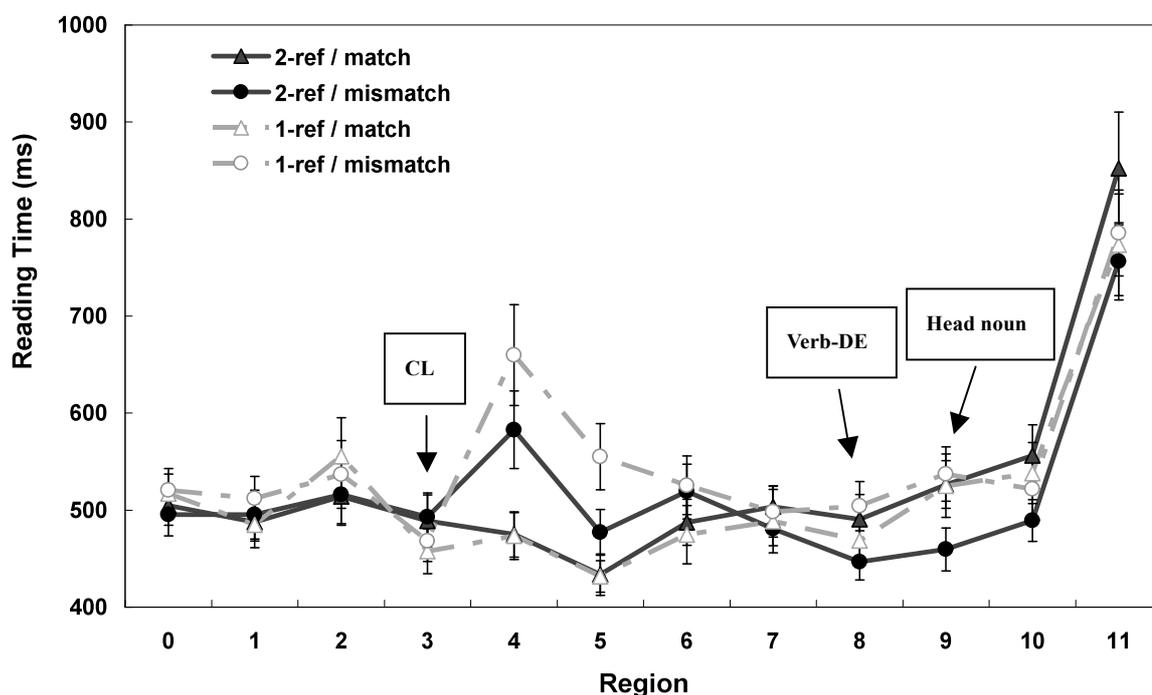
Context sentences$_0$
Zhege$_1$ xueqin-zhong$_2$, na-**liang/wei**$_3$ daxuesheng$_4$ chang-yi$_5$
This semester-during that-CL$_{vehicle/human}$ college-student often-with
xixinde$_6$ taidu$_7$ baoyang-de$_8$ motuoche$_9$ jingran$_{10}$ bei-tou-le$_{11}$.
careful attitude maintain-DE motorcycle suddenly was-stolen-ASP
'During this semester, the motorcycle that the college student often carefully maintains was suddenly stolen.'

Figure 3. Mean raw reading times (in ms) per region in Experiment 3c

Through Region 3 there were no main effects or interactions, except that at Region 0 (context sentence), the RT was longer in the 1-ref context conditions (519ms) than in the 2-referent context conditions (500ms). This difference was reliable only in the participants analysis ($F_1(1,27) = 10.20$, $p < .01$; $F_2(1,19) = .878$, $p < .36$). At the noun in Region 4 and at the following two regions there was a main effect of the classifier-match factor, due to the slower reading times in the classifier-mismatch conditions (*Region 4*: classifier-match = 474ms, classifier-mismatch = 622ms; $F_1(1,27) = 26.43$, $p < .001$; $F_2(1,19) = 34.24$, $p < .001$; *Region 5*: classifier-match = 433ms, classifier-mismatch = 516ms; $F_1(1,27) = 22.97$, $p < .001$; $F_2(1,19) = 18.85$, $p < .001$; *Region 6*: classifier-match = 481ms, classifier-mismatch = 523ms; $F_1(1,27) = 7.10$, $p < .05$; $F_2(1,19) = 8.22$, $p = .01$). Planned paired *t*-test comparisons confirmed that the main effect reflected significant slowdowns for mismatching classifiers at each level of the context factor in Regions 4-5 (2-referent context: *Region 4*: classifier-match = 474ms, classifier-mismatch = 582ms; $t_1(27) = 3.32$, $p = .002$; $t_2(19) = 2.75$, $p = .013$; *Region 5*: classifier-match = 433ms, classifier-mismatch = 477ms; $t_1(19) = 2.386$, $p = .024$; $t_2(19) = 1.376$, $p = 0.09$ (one-tailed); 1-referent context: *Region 4*: classifier-match = 473ms, classifier-mismatch = 660ms; $t_1(27) = 4.78$, $p < .001$; $t_2(19) = 3.53$, $p < .01$; *Region 5*: classifier-match = 431ms, classifier-mismatch = 555ms; $t_1(27) = 4.57$, $p < .000$; $t_2(19) = 3.16$, $p < .01$). Furthermore, at Regions 4-5 the particularly long reading times in the 1-referent/mismatch condition led to a main effect of context that was marginally significant in the participants analysis but not in the items analysis (*Region 4*: 2-referent context = 529ms, 1-referent context = 567ms; $F_1(1,27) = 3.74$, $p = .06$; $F_2(1,19) = 1.05$, $p = .318$; *Region 5*: 2-referent context = 455ms, 1-referent context = 494ms; $F_1(1,27) = 5.34$, $p < .05$; $F_2(1,19) = 2,83$, $p < .11$), and a significant context × classifier-match interaction that was marginally significant in the participants analysis but not in the items analysis (Region 4: $F_1(1,27) = 3.51$, $p = .07$; $F_2(1,19) = 1.01$, $p = .33$; Region 5: $F_1(1,27) = 7.00$, $p < .05$; $F_2(1,19) = 1.64$, $p = .22$). There were no main effects or interactions at the adverbial phrase in Region 7. The lack of effects in Region 7 is important, since this is the final region before the critical disambiguating regions at the end of the relative clause, and this suggests that any effects at the end of the relative clause were not confounded by continuing disruption from encountering the classifier-noun mismatch.

The critical regions were the embedded verb and RC marker *-de* in Region 8 and the head noun in Region 9. At Region 8 there was no main effect of context or classifier type. The interaction between context and classifier type was significant in the participants analysis but not in the items analysis (2-referent/match = 491ms, 2-referent/mismatch = 447ms, 1-referent/match = 468ms, 1-referent/ mismatch = 504ms; $F_1(1,27) = 7.17$, $p < .05$; $F_2(1,19) = 2.26$, $p = .15$). At Region 9, the main effect of context was significant in the participants analysis but not in the items analysis (2-referent context = 493ms, 1-referent context = 531ms; $F_1(1,27) = 7.87$, $p < .01$, $F_2(1,19) = 2.33$, $p = .14$), and the main effect of classifier type was marginally significant in the participants analysis but not in the items analysis (classifier-match = 526ms, classifier-mismatch = 499ms; $F_1(1,27) = 3.42$, $p = .076$, $F_2(1,19) = 2.51$, $p = .13$). Similarly, the interaction between context and classifier type was significant in the participants analysis but not in the items analysis (2-referent/match = 526ms, 2-referent/mismatch = 460ms, 1-referent/match = 525ms, 1-referent/mismatch = 537ms; $F_1(1,27) = 6.43$, $p < .05$, $F_2(1, 19) = 2.28$, $p < .15$). The contrast between the participants and items analyses at these regions could reflect the general variability in RTs across items. Thus, an additional 2 × 2 ANOVA was conducted using residual reading times instead of raw reading times at Regions 8-9. Since the residual RTs took word length into consideration and were based on each participant's average, the analysis of residual reading times reduced the variability that was related to word length for each participant. The results showed an overall pattern that was the same as the raw reading times. At Region 9, but not at Region 8, the main effect of context was found to be significant (2-referent context = -28ms, 1-referent context = 13ms; $F_1(1,27) = 8.76$, $p < .01$; $F_2(1,19) = 6.09$, $p < .05$), and the main effect of classifier type was not significant at either region. At both regions, the interaction between context and classifier type was significant in both the participants and the items analyses (*Region 8*: 2-referent/match = -50ms, 2-referent/mismatch = -113ms, 1-referent/match = -92ms, 1-referent/mismatch = -50ms; $F_1(1,27) = 7.39$, $p < .05$; $F_2(1,19) = 11.77$, $p < .01$; *Region 9*: 2-referent/match = 8ms, 2-referent/mismatch = -64ms, 1-referent/match = 3ms, 1-referent/mismatch = 22ms; $F_1(1,27) = 7.37$, $p < .05$; $F_2(1,19) = 7.68$, $p < .05$).

Additionally, planned comparisons on residual RTs tested the effect of classifier type within each level of the context factor at Regions 8-9. There was no significant effect of classifier type at these regions in the 1-referent context conditions (*Region 8*: $t_1(27) = 1.77$, *p*

< .09; $t_2(19) = 1.38$, $p = .185$; *Region 9*: ts < *1*), but in the 2-referent context conditions reading times were significantly shorter in the classifier-mismatch condition than in the classifier-match condition (*Region 8*: $t_1(27) = 2.46$, $p < .05$; $t_2(19) = 2.42$, $p < .05$; *Region 9*: $t_1(27) = 3.73$, $p < .001$; $t_2(19) = 3.35$, $p < .01$). Importantly, this RT pattern in the 2-referent conditions was the reverse of the RT pattern observed following the mismatching classifier-noun sequence in Regions 4-6. At the post-critical regions, there were no significant RT differences, except that at Region 10 there was a significant main effect of classifier type, with faster RTs in the classifier-mismatch condition (547ms) than in the classifier-match condition (530ms) ($F_1(1,27) = 5.12$, $p < .05$; $F_2(1,19) = 6.31$, $p < .05$).

Discussion

The above results yield two clear patterns. First, as expected, the classifier-noun mismatch sequence induced processing difficulty, as reflected by the slowdown found in both the 1-referent and the 2-referent context conditions. Second, the reversal pattern in reading times at the RC disambiguating regions was found only in the 2-referent context conditions. This shows that only in the 2-referent context is the recognition of an RC structure facilitated by the classifier-noun mismatch cue. In other words, when context creates referential ambiguity due to multiple referents of the same type, an RC interpretation is successfully generated from the classifier-noun mismatch cue. On the other hand, when context contains only one potential referent for the NP, thereby creating no need for restrictive modification to fulfill presuppositional requirements, an RC interpretation is not generated from the classifier-noun mismatch cue. This finding suggests that the parser is able to exploit the classifier-noun mismatch cue to recognize an RC structure only in a RC-felicitous context, but not in a non-RC-felicitous context. We take this contrast as an important piece of evidence that discourse information impacts the parser's ability to use the indirect cue of a classifier-noun mismatch sequence to recognize an upcoming RC structure.

Critically, these results suggest that in an on-line setting an RC-supportive context is necessary to help the parser to generate a grammatical RC structure based on the indirect cue of a classifier-noun mismatch sequence. There is, however, a possible alternative explanation that prevents us from firmly concluding that it is the classifier-noun mismatch cue that leads the parser to generate the RC structure in the 2-referent context. In this experiment, the

classifier-noun mismatch sequence, such as in *na-liang daxuesheng… "that-CL_{vehicle}* *college-student…",* contained the demonstrative pronoun *na* "that", which is definite and singular. This means that already at the point of *na-liang "that-CL_{vehicle}",* comprehenders may have noticed the problem of referential ambiguity created by the 2-referent context, and hence may have tried to resolve the ambiguity immediately by anticipating a restrictive modifier. Thus, it is possible that the generation of the RC structure reflects the *demonstrative(singular)-classifier* sequence rather than the classifier-noun mismatch sequence. If the parser starts to anticipate a RC structure at the point of the demonstrative(singular)-classifier, it would suggest that context is used in selecting a possible structure from among the structural alternatives that are available at this point, rather than helping the parser to generate the correct structure based on the classifier-noun mismatch sequence. In order to address this concern, Experiment 3d was designed to remove the confound caused by the singular demonstrative, in order to examine whether the parser can generate the RC structure based on the classifier-noun mismatch sequence in a supportive context.

Experiment 3d: On-line Reading Experiment

Experiment 3d examined whether a classifier-noun mismatch was sufficient to trigger generation of an RC structure online, in a context that was consistent with a RC. We tested this by using an indefinite determiner rather than a definite determiner in the classifier-noun pair. The indefinite lacks the uniqueness presupposition associated with a definite NP, and hence the indefinite noun is compatible with a 2-referent context, with no need for an additional restrictive modifier. If the combination of the supportive context and the classifier-noun mismatch is still sufficient to cue the generation of a relative clause, then we can more confidently conclude that context affects structure generation.

The experimental stimuli were similar to Experiment 3c. Each target sentence was preceded by a context sentence that established multiple potential referents for the determiner in the target sentence. In addition, the particle *suo*, which provides an early direct cue for head-final RCs (Experiment 1), was used in the control conditions. One set of test conditions is illustrated in (17-18).

(17) Context sentence

Zai zhe-jia   chongwu yiyuanli, you   ji-wei   shouyishi fuze

in this-CL$_{building}$ pet   hospital have several-CL$_{human}$ vet   responsible

zhaogu *ji-zhi*   shengbingde xiaogou.

care   several-CL$_{animal}$ sick   dog

'In this animal hospital, there are several vets responsible for taking care
  of several sick dogs.'

(18) a. *Classifier-match Condition*

Zuotian   wanshang, you yi-***wei***   shouyishi yi zhuanyede fangshi (suo)

yesterday   night YOU one-CL$_{human}$ vet   with professional way (SUO)

xixindi   zhiliao de yi-<u>zhi</u>   xiaogou yijing   quanyu-le.

carefully   treat   DE one-CL$_{animal}$ dog   already recover-LE

'Last night, a dog which one of the vets treated carefully in a professional way had
already recovered.'

    b. *Classifier-mismatch Condition*

Zuotian   wanshang, you yi-***zhi***   shouyishi yi   zhuanyede fangshi (suo)

yesterday night   YOU one-CL$_{animal}$ vet   with professional way (SUO)

xixindi   zhiliao de xiaogou yijing quanyu-le.

carefully   treat   DE dog   already recover-LE

'Last night, one of the dogs which a vet treated carefully in a professional way had
already recovered.'

The context sentence in (17) contains multiple referents associated with the human classifier *wei* and the animal classifier *zhi*. In the test sentences, the word *you* that appeared before the indefinite determiner is interpreted as 'there is one of the'. In the target test sentences, as illustrated in (18), existential sentences were used. These sentences require indefinite subject NPs and are introduced by the particle *you*. In the classifier-match condition in (18a), the first classifier matched the immediately following noun, *shouyishi* 'vet'. Note that in this condition, another determiner, *yi-zhi*, 'one-CL$_{animal}$', was added immediately before the RC head noun. This determiner was inserted to ensure that the following noun *xiaogou* "dog" would have an indefinite reading, as required by the *you*-existential sentences. The sentence would be ungrammatical if this additional determiner were not present. The presence of this additional determiner in the classifier-match condition was unavoidable, but it made it more difficult to directly compare the critical regions (i.e. the RC head noun region) in the classifier-match conditions and the classifier-mismatch

conditions. We therefore focused on the effect of the presence of the direct RC cue *suo* within each level of the classifier match factor. We reasoned that if the parser has already correctly identified the presence of an RC before it encounters the particle *suo*, then *suo* should confer no additional benefit in the processing of RCs. On the other hand, if the parser's first effective cue for an RC is the particle *suo*, then we should expect to find the same reading time benefit observed in Experiment 1b. We predicted that if the RC-supporting context and the classifier mismatch together allow the parser to generate an RC structure, then we should find no additional RT benefit from the particle *suo* in the classifier-mismatch conditions.

Methods

Participants. Twenty-eight native speakers of Mandarin Chinese spoken in Taiwan participated in the experiment. They were undergraduate students from National Tsing Hua University and National Chao Tung University in Taiwan. They were naïve to the purpose of the experiment. The experiment lasted about 30 minutes.

Materials and Design. This experiment used a 2 × 2 factorial design that manipulated the presence of the particle *suo* (*suo* vs. no *suo*) and the classifier match (classifier-match vs. classifier-mismatch). However, the critical comparisons were within each level of the classifier match factor, due to lexical differences between the matching and mismatching conditions at the critical regions. Twenty sets of four conditions were distributed among four lists in a Latin Square design. Each participant saw exactly one of the lists intermixed with eighty-four filler items in a random order. The procedure for this experiment was identical to the self-paced reading procedure used in the previous on-line experiments.

Data Analysis. Analyses were conducted on comprehension accuracy and reading times. All participants had comprehension accuracy scores higher than 75%, so none were excluded on this basis. Reading times for all trials for which the comprehension question was answered incorrectly were excluded. Reading times that exceeded a threshold of 2.5 standard deviations above a participant's mean RT for each region were replaced by the threshold value. In addition, one set of the test items was removed because one of its four versions was answered incorrectly by all participants. The means and analyses presented below are based on the remaining trials.

Results

Comprehension Accuracy. The average comprehension accuracy across all items was 90%. The average comprehension accuracy for the test conditions was 89%. The average comprehension accuracy for the individual conditions was 91% in the classifier-match/no-*suo* condition, 85% in the classifier-mismatch/no-*suo* condition, 91% in the classifier-match/*suo* condition, and 89% in the classifier-mismatch/*suo* condition. There were no main effects or interactions in the analysis of the accuracy scores (Fs < 1).

Reading times. Reading times at each region were first analyzed by a repeated-measures 2 × 2 ANOVA, with the factors classifier type (match/mismatch) and the presence of *suo* (no-*suo*/*suo*) as within subjects factors. Data from twenty-eight participants were considered in the analysis. The average reading times for each region and condition are shown in Figure 4. The only main effects or interactions in this overall ANOVA were found in Regions 6-8 following the classifier-noun sequence and in Region 10 at the adverb following the particle *suo*. At Regions 6-8 reading times were significantly longer in the classifier-mismatch conditions than in the classifier-match conditions (*Region 6*: classifier-match = 384ms, classifier-mismatch = 462ms; $F_1(1,27) = 19.94$, $p < .001$, $F_2(1,18) = 13.26$, $p < .01$; *Region 7*: classifier-match = 396ms, classifier-mismatch = 455ms; $F_1(1,27) = 3.24$, $p < .05$, $F_2(1,18) = 12.83$, $p < .01$). At the particle *suo* in Region 9 RTs were longer in the classifier-mismatch condition than in the classifier-match condition, but the effect was only marginally significant (classifier-match = 361ms, classifier-mismatch = 413ms; $F_1(1,27) = 2.97$, $p < .1$, $F_2(1,18) = 4.18$, $p < .06$). At the adverb in Region 10 RTs were longer in the *suo* conditions than the no-*suo* conditions, but the effect was only marginally significant (*suo* = 439ms, no-*suo* = 401ms; $F_1(1,27) = 3.27$, $p < .09$, $F_2(1,18) = 2.85$, $p < .11$).
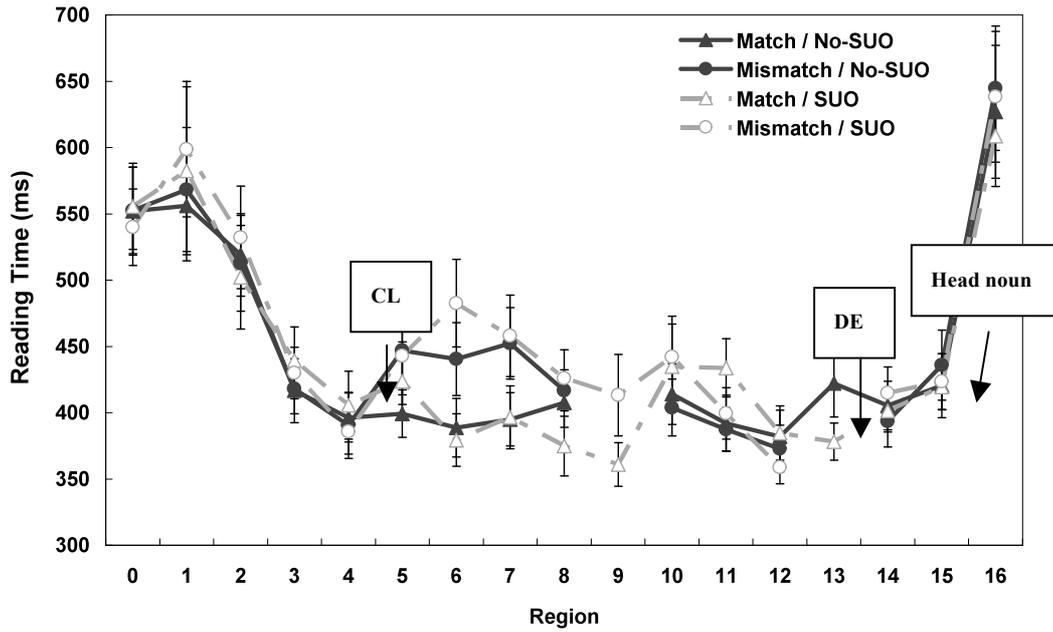
Context sentence₀ (with multiple identical referents)

Zuotian₁ wanshang₂, you₃ yi-***wei/chi***₄ shouyishi₅ yi₆ zhuanyede₇ fangshi₈ (suo)₉
yesterday night YOU one-CL_{human/animal} vet with professional way (SUO)

xixindi₁₀ zhiliao₁₁ de₁₂ (yi-zhi₁₃) xiaogou₁₄ yijing₁₅ quanyu-le₁₆.
carefully treat DE one-CL_{animal} dog already recover-LE

'Last night, a dog which one of the vets treated carefully in a professional way had already
  recovered.'

Figure 4: Mean reaction times (ms) per region in Experiment 3d

Due to lexical differences between conditions, analyses in the critical regions at the end of
the RC (Regions 12-14) separately tested for effects of the particle *suo* within each level of
the classifier match factor. In the classifier-mismatch conditions there were no significant
differences. In the classifier-match conditions RTs differed at the determiner in Region 13,
immediately following the RC marker *de*. Reading times were longer in the no-*suo* condition
than in the *suo* condition, although the difference was only marginally significant in the
participants analysis (no-*suo* = 422ms, *suo* = 378ms; $F_1(1,27) = 3.17$, $p < .09$, $F_2 (1,18) =$
7.62, $p < .05$). This contrast suggests that the presence of the RC cue *suo* benefited RC
processing only in the absence of the classifier-mismatch cue.

Discussion

This experiment yielded two results of interest. First, the previous finding that the

particle *suo* can facilitate the processing of head-final RCs was replicated, as evidenced by the faster reading times at Region 13 in the classifier-match/*suo* condition than in the classifier-match/no-*suo* condition. This reinforces our claim that the parser is able to use the direct cue from the particle *suo* to detect an upcoming RC and avoid a garden path.

Second, the fact that the particle *suo* conferred no corresponding processing benefit at the end of the RC in the classifier-mismatch conditions suggests that an indefinite classifier-mismatch sequence can combine with an RC-supporting context to cue RC generation. This conclusion is based on the assumption that if the classifier-mismatch cue was ineffective, then the particle *suo* would have yielded the same benefit in the classifier-mismatch conditions as it did in the classifier match conditions, contrary to fact.

The results of this experiment resolve the question about the locus of the benefit of the RC-supporting 2-referent context. If the early detection of the RC structure in Experiment 3c were due to a clash with the definite determiner rather than to the classifier-noun mismatch, then early detection of the RC structure should have failed with the indefinite determiner in Experiment 3d, where the classifier did not carry the presupposition that the referent should be "uniquely identifiable" to the reader (Gundel et al., 1993). Instead, Experiment 3d shows that simply the presence of two referents associated with the same classifier is sufficient to allow the parser to recognize the import of the classifier-mismatch cue. In summary, the indirect mismatching classifier cue is sufficient to trigger generation of an RC, but only when an RC is supported by the discourse.

General Discussion

Three sets of experiments are reported which investigate online structure generation in Chinese in response to indirect and direct syntactic cues as well as discourse cues. Experiments 1a-b established that unanticipated RCs in Chinese induce processing slowdown at the disambiguation point at the end of the RC, and that an early direct cue from the particle *suo* (indicating an RC) can reduce this downstream processing bottleneck. Experiments 2a-b examined an indirect-but-unambiguous cue for an RC, a semantic mismatch between a classifier and an adjacent noun. Results indicated that this cue can be utilized by readers in an offline task, but that the indirect cue alone was insufficient to prevent a garden path in an on-line reading task. Experiments 3a-d also focused on the classifier mismatch cue for an RC,

but in the presence of semantic contexts that support restrictive relative clause modifiers. Off-line and on-line tests suggested that the classifier mismatch is a more effective cue for an RC when it is presented in a supportive discourse context. Relative to control sentences that lacked the classifier mismatch cue, the classifier mismatch itself led to substantial processing disruption, but it had the opposite effect at the disambiguation point at the end of the RC, leading to reduced processing difficulty.

The motivation for investigating cues for Chinese head-final RCs was to better understand the role of semantic and discourse context on syntactic structure generation in language processing. Many previous studies have examined the effects of discourse context on structural ambiguity resolution, with numerous reports of early effects of context on how comprehenders resolve structural ambiguities (Altmann, Garnham, & Henstra, 1994; Altmann & Steedman, 1988; Chambers, Tanehaus, & Magnuson, 2004; Hoeks, Vonk, & Schriefers, 2002; Liversedge, Pickering, Branigan, & Van Gompel, 1998; Ni & Crain, 1990; Spivey-Knowlton & Tanenhaus, 1994; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; van Berkum, Brown, & Hagoort, 1999). Such findings are interesting, but they do not directly bear on the question of whether the language processor is an interactive or encapsulated cognitive system (Fodor, 1983), because they do not show that contextual information changes the internal operations of the syntactic processor. Effects of context on ambiguity resolution are consistent with a syntactic processor that generates parses without regard to context, but that submits its outputs for selection based on contextual and other information (Boland & Cutler, 1996). Some previous studies have examined the effects of context on the processing of unambiguous sentences (Grodner et al., 2005; van Berkum et al., 1999), but as discussed in the Introduction these studies do not show that structure generation operations in the syntactic processor are affected by context manipulations. The effects of contextual facilitation or disruption observed in these studies plausibly reflect non-syntactic processes.

The eight experiments described here collectively build an argument that context directly impact the structure generation process, and not just the structure selection process. The steps in the argument are as follows. We first show that Chinese head-final RCs lead to (mild) processing difficulty, plausibly because they can initially be analyzed as simple main

clauses or complement clauses until the RC-final particle –*de* disambiguates and forces reanalysis. As evidence for this difficulty, we show that the presence of a direct RC cue, the particle *suo*, leads to faster reading times at the end of the RC (Experiment 1). The next step in the argument is to show that the parser fails to successfully analyze a different type of RC cue, the indirect cue provided by a mismatching classifier-noun sequence (Experiment 2). This provides a situation where the syntactic processor initially outputs zero grammatical analyses of the classifier mismatch. Finally, we show that addition of a supportive discourse context creates an environment where the parser becomes able to recognize the classifier mismatch as an RC cue (Experiment 3). In other words, the context takes a syntactic processor that would otherwise be unable to generate any grammatical analyses for the classifier mismatch and makes it able to find a grammatical analysis. We take this to show that context can alter the internal operations of the syntactic processor, rather than merely having access to the outputs of syntactic parsing.

A possible objection to this argument involves the second step: it might not be true that the parser initially finds zero grammatical analyses of the classifier mismatch sequence. This is because Chinese allows noun ellipsis constructions in which a classifier appears without an associated overt noun. Consequently, the classifier-noun mismatch sequence might be successfully analyzed as a sequence of two independent NPs, as in (9b) above. If the parser generates this analysis, then it becomes possible that the parser's analysis of classifier mismatch sequences always consists of the noun ellipsis structure and an RC, and that the role of context is merely to alter the structure selection process, thereby neutralizing the argument for interactivity in the syntactic processor. The results of the sentence fragment completion study (Experiment 2a) suggest that this concern does not apply to our studies. If the noun ellipsis structure were commonly chosen as an analysis of classifier mismatch in the on-line study, then we should have expected to find many examples of such structures in the fragment completion study. In fact, such completions were extremely rare. They accounted for only 6% of completions overall, and aside from 3 (out of 36) participants who frequently used this structure in their completions, other participants used the noun ellipsis structure in only 3% of completions. The rare use of this completion may be a consequence of the discourse conditions on the use of the noun ellipsis construction in Chinese.

Our finding that Chinese speakers are not able to use the classifier mismatch cue for

RCs online, unless a supportive discourse context is provided, contrasts with findings from related tests in Japanese, where reading time studies similar to our Experiment 2b show that a classifier mismatch eases processing at the end of a RC, even in the absence of a supportive context (Yoshida, 2006). There are a number of possible reasons why the classifier mismatch cue may be more effective for Japanese speakers. First, Japanese has case marking on nominals, which may provide a richer set of morphosyntactic cues that can be combined with the classifier mismatch information. Second, Japanese is strongly head-final, whereas Chinese is head-final only in noun phrases, and this difference may make Japanese speakers more efficient in the use of cues to head-final structures. Third, there is some flexibility in the position of the classifier in Chinese, such that it may appear following the RC, raising the possibility that the word order of the classifier-noun mismatch sequence is less natural for Chinese speakers (Wu, Haskell, & Andersen, 2006), making the cue less effective than its Japanese counterpart. The current results do not allow us to choose among these or other accounts of the cross-language difference. For purposes of the main argument here it is not critical to determine exactly why Chinese speakers fail to exploit the classifier mismatch cue online. The important point for the current argument is the fact that Chinese speakers generate no grammatical analyses of the classifier mismatch cue in the absence of supportive context.

Turning to the step of the argument that involves effects of discourse context, we argued that the effect of an RC-supporting discourse context is to change the classifier mismatch cue from an ineffective cue into a cue that successfully signals an RC structure online (Experiment 3c). The argument for effects of context on structure generation crucially depends on the claim that the supportive discourse context and the classifier mismatch cue are jointly responsible for the facilitation in RC processing. If a 2-referent context alone were sufficient to facilitate RC processing, then both the classifier-match and the classifier-mismatch conditions should have shown facilitation at the end of the RC in Experiment 3c. However, it was only the 2-referent/mismatch condition that showed faster RTs at the end of the RC, indicating that both the supportive context and the indirect cue were needed to facilitate RC processing. Relatedly, it is important to address the possibility that it is the clash between a 2-referent discourse context and a singular definite determiner that leads the parser to consider an RC structure, and not the effect of the classifier mismatch

itself. We tested this possibility in Experiment 3d by combining classifiers with indefinite determiners. The finding that the particle *suo* conferred no additional benefit for RC processing in sentences with a supportive context and a classifier mismatch suggests that it is the classifier mismatch rather than the definite determiner that is crucial to facilitate RC processing.

A different challenge to our argument involves the possibility that discourse context only affects structure selection, rather than directly impacting the range of structural analyses that the syntactic processor generates. Our argument for effects of context on structure generation relied on the claim that, in the absence of a supportive discourse, the parser arrives at no grammatical analyses of the classifier mismatch sequence. We therefore reasoned that it would be difficult to attribute the parser's failure to use the classifier mismatch cue to a preference for an alternative parse. However, it could be objected that we restricted our attention to grammatically well-formed parses. If we include the possibility that the syntactic processor, when presented with a classifier mismatch sequence, also generates a grammatically inappropriate non-RC parse, then it might be possible to maintain that the syntactic processor generates the same outputs irrespective of context, and that the effect of a context is simply to change the ranking of alternatives at the structure selection stage. Under this view, the syntactic processor always analyzes a classifier mismatch sequence as ambiguous, generating a grammatical RC parse and an ungrammatical non-RC parse. In the absence of a supportive discourse context, the structure selection process would then treat the grammatical RC parse as so infelicitous that the ungrammatical non-RC parse is favored instead. The effect of a supportive context would be to shift the ranking of the two parses, rather than changing the range of alternatives that the syntactic processor generates.

The logic of this non-interactive account of our findings is sound, but we consider it to be unlikely. First, it requires the assumption that in Experiment 2b the classifier mismatch cue always led to generation of an RC structure, despite the fact that reading times showed no evidence of a benefit of early RC detection. Second, the account relies on the inaccurate assumption that the RC parse is infelicitous in the absence of a multiple-referent context. Although it is true that multiple-referent contexts make modification with a restrictive RC particularly natural, it is not the case that modification with a RC is infelicitous in the absence of a prior multiple-referent context. Third, the account assumes that grammatical

well-formedness has so little value for the structure selection process that a well-formed parse can be ignored in favor of a structurally simple yet ungrammatical parse. This assumption is at odds with much evidence for the effectiveness of grammatical cues in language processing, including the success of the unambiguous control conditions in very many studies of structural ambiguity resolution.

An alternative that we consider more likely is that Chinese speakers in our studies did indeed often generate an ungrammatical simple clause analysis of the classifier mismatch sequence, but that they did so as a last resort, only after failing to find a grammatical parse. This is consistent with the evidence of many simple clause completions in the sentence fragment completion study (Experiment 2a), and with the lack of processing benefit from the classifier mismatch in the corresponding reading time study (Experiment 2b).

We should note that our arguments in this article address only one of the two main considerations that Fodor (1983) raised in favor of a modular, informationally encapsulated language processing system. The first consideration is the empirical claim that contextual information does not alter the internal workings of the parser. Our experimental evidence challenges this empirical claim, and we have argued that the Chinese evidence cannot readily be explained away in terms of selection processes. Fodor's second consideration involves a more theoretical challenge. Fodor pointed out that a mental module can only act upon information that it is able to interpret, and argued that the representations manipulated by different processing systems are of different types, and hence are not mutually interpretable. Applied to the current case, this amounts to the observation that it is unclear how semantic information about the similarity of referents in a discourse model should lead to generation of syntactic relative clause structure. If the syntactic processor independently generates an RC structure, it is straightforward to capture the fact that this structure is compatible with a multi-referent discourse structure. But it is far less clear how the discourse structure could itself lead to generation of the RC structure. Our studies do not resolve this important problem.

We have emphasized the difference between structure generation processes and structure selection processes, arguing that it is the structure generation processes that are most relevant to modularity arguments. However, much recent work in sentence processing under the broad heading of constraint-based lexicalism has attempted to neutralize Fodor's theoretical

argument by eschewing the distinction between generation and selection. If it is possible to encode different types of evidence for a parse (syntactic, semantic, pragmatic, etc.) using a common vocabulary, and if the representation of the parse itself can be directly linked to the representation of the evidence for the parse, then Fodor's second, theoretical argument largely disappears. In constraint-based lexicalist models it is common to assume that individual parsing choices are represented by units whose activation encodes the strength of the evidence for that parse. This encoding is typically understood as analogous to similar views of lexical activation. Different sources of evidence for the parse are all encoded using a common vocabulary, in terms of the activation values of other nodes in a neural network. Meanwhile, the sources of evidence for a parse and the parse itself are linked via direct connections that are associated to one another with continuously varying weights (MacDonald, 1993; Spivey & Tanenhaus, 1998; Trueswell & Tanenhaus, 1994). In this conception of the problem, it is straightforward to integrate varying sources of evidence to vary the strength of support for a given parse.

To the extent that the constraint-based lexicalist approach is successful in eliminating the distinction between generation and selection that we have emphasized here, it undermines our primary goal of assessing the impact of contextual information on structure generation processes. If parses are merely activated in memory, rather than explicitly constructed via assembly of words and phrases, then there are no real structure generation processes *per se*, and hence it makes little sense to ask whether the internal operations of the structure generation system have access to information from discourse and other sources. However, we contend that it is not possible to eliminate the need for structure generation processes by emphasizing the importance of lexical information in parsing. In the case of Chinese classifier mismatches, for example, the relative clause structure is not a part of the lexical representation of the classifier, nor is it a part of the lexical representation of the following noun (which is ultimately parsed as the subject of the RC), and nor do the classifier and mismatching noun enter into a direct lexical relation with one another. It is a theorem of the grammar of Chinese that the mismatching classifier-noun sequence can be parsed as the left edge of a RC structure, but this fact cannot be retrieved by simply activating the long-term lexical representations of the classifier and the noun. Parsing and combinatorial operations are required in order to generate the structure. Consequently, it is meaningful to ask whether

and how those structure generation operations are influenced by extra-grammatical information. The Chinese case is particularly interesting, as recognizing the indirect RC cue so clearly requires more than activation of the lexical items in the input string, but similar arguments on the feasibility of a full lexical approach could be made based on many other phenomena (Frazier, 1995; Sturt & Lombardo, 2005). Our finding that contextual information impacts the effectiveness of the classifier mismatch cue therefore goes beyond previous evidence for effects of context on structure selection in parsing.

In sum, this study examined a case of parsing failure in Chinese where the parser is not able to use an indirect cue to generate the appropriate upcoming structure without an appropriate context. This breakdown is not related to structural complexity or memory overload, but instead involves a cue that is not correctly used. Since no ambiguity is involved in the classifier-noun mismatch sequence, the role of context in this case is not to help resolve an ambiguity. The effect of adding contextual information is to help the parser to overcome its failure to recognize a structural cue. We therefore conclude that structural and non-structural information can be combined specifically at the generation stage of parsing, even to identify unambiguous parses. We suggest that these findings present stronger evidence for the non-encapsulation of structure generation processes than has been possible in previous studies that have focused on structural ambiguity resolution.

References

Abney, S. P., & Johnson, M. (1991). Memory requirements and local ambiguity of parsing strategies. *Journal of Psycholinguistic Research, 20*(3), 233-250.

Altmann, G., Garnham, A., & Dennis, Y. (1992). Avoiding the garden-path: Eye movements in context. *Journal of Memory and Language, 31*, 685-712.

Altmann, G., Garnham, A., & Henstra, J.-A. (1994). Effects of syntax in human sentence parsing: Evidence against a structure-based proposal mechanism, *Journal of Experimental Psychology: Learning, Memory and Cognition* (Vol. 20, pp. 209-216).

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 77*, 247-264.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191-238.

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

Aoshima, S., Yoshida, M., & Phillips, C. (2009). Incremental processing of coreference and binding in Japanese. *Syntax, 12*, 93-134.

Becker, C. A. (1976). Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 556-566.

Boland, E. J., & Cutler, A. (1996). Interaction with autonomy: Multiple output models and the inadequacy of the Great Divide. *Cognition, 58*, 309-320.

Bornkessel, I., Schlesewsky, M., & Friederici, A. (2003). Contextual information modulates initial processes of syntactic integration: The role of inter-versus intrasentential predictions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 871-882.

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 813-840.

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases, *Journal of Memory and Language, 33*, 251-283

Britt, M. A., Garrod, S. & Rayner, K. (1992). Parsing and discourse: Context effects and their limits, *Journal of Memory and Language*, 31, 293-314

Brown, C. M., Hagoort, P., & Vonk, W. (1997). *Semantic effects on syntactic analyses: ERP evidence*. Paper presented at the Poster presented at the Fourth Annual Meeting of the Cognitive Neuroscience Society.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 687-696.

Chen, P. (1996). Pragmatic interpretations of structure topics and relativization in Chinese. *Journal of Pragmatics, 26*, 389-406.

Chiu, B. (1993). *The inflectional structure of Mandarin Chinese.* Unpublished Ph.D. Dissertation, University of California, Los Angeles, Los Angeles, CA.

Chiu, B. (1995). An object clitic project in Mandarin Chinese. *Journal of East Asia Linguistics, 4*(2), 77-117.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen & A. Zwicky (Eds.), *Natural Language Processing: Psychological, Computational and Theoretical Perspectives* (pp. 320-358). Cambridge: Cambridge University Press.

Farmer, T. A., Anderson, S. E., & Spivey, M. J. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language, 57*, 570-595.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348-368.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Frazier, L. (1987). Theories of syntactic processing. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural language processing* (pp. 291-307). Cambridge, MA: MIT Press.

Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research, 24*, 437-468.

Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language, 26*, 505-526.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition, 68*, 1-76.

Gibson, E. (2000). The dependency locality theory: a distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz & W. O'Neil (Eds.), *Image, language, brain* (pp. 95-126). Cambridge, MA: MIT Press.

Grice, P. (1957). Meaning. *The Philosophical Review, 66*, 377-388.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input. *Cognitive Science, 29*, 261-290.

Grodner, D., Gibson, E., & Watson, D. (2005). The influence of contextual contrast on syntactic processing: Evidence for strong-interaction in sentence comprehension. *Cognition, 95*, 275-296.

Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 274-307.

Hoeks, J. C. J., Vonk, W., & Schriefers, H. (2002). Processing coordinated structures in context: The effect of topic-structure on ambiguity resolution. *Journal of Memory and Language, 46*, 99-119.

Inoue, A., & Fodor, J. (1995). Information-paced parsing of Japanese. In R. Mazuka & N.

Nagai (Eds.), *Japanese Sentence Processing* (pp. 9-36). Hillsdale, NJ.

Johnson-Laird, P. N. (1983). *Mental Models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, Massachusetts: Harvard University Press.

Just, M. A., Carpenter, P. A., & Wooley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 3*, 228-238.

Kaan, E., & Swaab, T. Y. (2003). Electrophysiological evidence for serial sentence processing: A comparison between non-preferred and ungrammatical continuations. *Cognitive Brain Research, 17*, 621-635.

Kaiser, E., & Trueswell, J. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition, 94*, 113-147.

Kamide, Y., & Mitchell, D. C. (1999). Incremental pre-head attachment in Japanese parsing. *Language and Cognitive Processes, 14*, 631-662.

Lombardo, V., & Sturt, P. (2002). Incrementality and lexicalism: A treebank study. In S. Stevenson & P. Merlo (Eds.), *Lexical Representations in Sentence Processing* (pp. 137-155): John Benjamins

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition, 25*, 71-102.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology, 10*, 29-63.

Mazuka, R., & Itoh, K. (1995). Can Japanese speakers be led down the garden path. In R. Mazuka & N. Nagai (Eds.), *Japanese sentence processing* (pp. 295–329): Lawrence Erlbaum.

McClelland, J. L., & Elman, J. L. (1986). The TRACE Model of Speech Perception. *Cognitive Psychology, 18*, 1-86.

McDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676-703.

Mitchell, D. C., Corley, M. M. B., & Garnham, A. (1992). Effects of context in human sentence parsing: Evidence against a discourse-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 69-88.

Miyamoto, E. T. (2002). Case markers as clause boundary inducers in Japanese, *Journal of Psycholinguistic Research* (Vol. 31, pp. 307-347).

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*, 165-178.

Murray, W. S., & Liversedge, S. P. (1994). Referential context effects on syntactic processing. In C. Clifton, Jr. , L. Frazier, K. Rayner & C. Clifton, Jr. (Eds.), *Perspectives on sentence processing* (pp. 359–388): Lawrence Erlbaum.

Ni, W., & Crain, S. (1990). *How to resolve structural ambiguities*. Paper presented at the Proceedings of the North East Linguistics Society.

Norris, D. (1986). Word recognition: context effects without priming. *Cognition, 22*, 93-136.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189-234.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 786-803.

Pulvermüller, F., Assadollahi, R., & Elbert, T. (2001). Neuromagnetic evidence for early semantic access in word recognition. *European Journal of Neuroscience, 13*, 201-205.

Rayner, K., Garrod, S. C., & Perfetti, C. A. (1992). Discourse influences during parsing are delayed. *Cognition, 45*, 109-139.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71*, 109-147.

Sekerina, I. A. (2003). Scrambling and processing: dependencies, complexity, and constraints. In S. Karimi (Ed.), *Word order and scrambling* (pp. 301–324). Malden, MA: Blackwell Publishers.

Sereno, S., Brewer, C., & O'Donnell, P. (2003). Context effects in word recognition: Evidence for early interactive processing. *Psychological Science, 14*, 328-333.

Simpson, G. B. (1984). Lexical ambiguity and its role in models of word recognition. *Psychological Bulletin, 96*, 316-340.

Smith, M. C., & Besner, D. (2001). Modulating semantic feedback in visual word recognition. *Psychonomic Bulletin & Review, 8*, 111-117.

Spivey-Knowlton, M. J., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition, 55*, 227-267.

Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1994). Referential context and syntactic ambiguity resolution. In L. Frazier & K. Rayner (Eds.), *Perspectives on Sentence Processing*. Hillsdale, NJ: Erlbaum.

Spivey-Knowlton, M. J., Trueswell, J., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology, 47*, 276-309.

Spivey, M., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1521-1543.

Spivey, M. J., Tanehaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution *Cognitive Psychology, 45*, 447-481.

Stabler, E. P. (1994). The finite connectivity of linguistic structure. In C. Clifton, L. Frazier & K. Rayner (Eds.), *Perspectives on Sentence Processing* (pp. 303-336).   Hillsdale,

NJ: Erlbaum.

Sturt, P., & Crocker, M., W. (1996). Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes, 11*, 449-494.

Sturt, P., & Lombardo, V. (2005). Processing coordinated structures: Incrementality and connectedness. *Cognitive Science: A Multidisciplinary Journal, 29*, 291-305.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence Comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of Perception and Cognition: Speech, Language, and Communication* (Vol. 11, pp. 217-262). San Diego: San Diego: Academic Press.

Ting, J. (2003). The nature of the particle suo. *Journal of East Asian Linguistics, 12*, 121-139.

van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*, 147-182.

Van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In A. Kennedy, R. Radach, H. Dieter & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 621–648): North Holland.

Wu, F., Haskell, T., & Andersen, E. (2006). *The interaction of lexical, syntactic, and discourse factors in on-line Chinese parsing: Evidence from eye-tracking*. Paper presented at the 19th Annual CUNY Conference on Human Sentence Processing.

Xiang, M., Munn, A., Schmitt, C., & Ferreira, F. (2003). *Garden path and null objects in Chinese sentence comprehension.* Paper presented at the 16th Annual CUNY conference on Human Sentence Processing, MIT, Cambridge, MA.

Yoshida, M. (2006). *Constraints and mechanisms in long-distance dependency formation.* Unpublished Ph.D. Dissertation, University of Maryland, College Park, MD.

Yoshida, M., Aoshima, S., & Phillips, C. (2004). Relative clause prediction in Japanese. Paper presented at the Seventeenth Annual CUNY conference on Human Sentence Processing.