Running Head: Reading span and plausibility

Effects of reading span and plausibility in the reanalysis of *wh*-gaps by Chinese-English L2 speakers (TO APPEAR IN *SECOND LANGUAGE RESEARCH*)

Paola E. Dussias *Penn State University* and **Pilar Piñar** *Gallaudet University*

Received July 2007; revised January 2008; accepted March 2008

Address for correspondence: Paola E. Dussias, Department of Spanish, Italian and Portuguese, 211 Burrowes Building, The Pennsylvania State University, University Park, PA 16802 USA
email: pdussias@psu.edu

Abstract

This study utilizes a moving window technique to investigate how individual cognitive resources (operationalized in terms of reading span scores) might modulate the extent to which native English speakers and Chinese second-language learners of English utilize plausibility information to recover from an initial misparse in the processing of long-distance *wh*-questions. Consistently with findings in the previous literature, both groups postulated a filler-gap dependency at the earliest possible position. This was evidenced by subject-object extraction parsing asymmetries that were indicative of an initially incorrect filler-gap analysis. Additionally, it was found that plausibility did not prevent initial misparses, but affected how fast subjects recovered from misanalysis, with implausible analyses facilitating recovery. However, only the English L-2 participants in the higher span group resembled English native readers in their ability to exploit plausibility information in this way. We conclude that the individual cognitive resources of the learner are an important factor in determining the extent to which sentence processing might be qualitatively similar or different in an L1 and an L2.

Keywords:

I Introduction

Increased interest in sentence processing by second language (L2) speakers has brought attention to the kind of information that is available to the parser during sentence comprehension in a second language and to the differences and similarities in the mechanisms that monolingual speakers and L2 speakers utilize. Some studies suggest that adult L2 speakers construct shallower syntactic representations in their second language than their native (L1) speaker counterparts and, therefore, do not use structural information in the same way as individuals reading in their L1 (e.g., Clahsen & Felser, 2006). On the other hand, other studies report that L2 readers do not seem to differ qualitatively from L1 readers in the way they use non-structural information, such as plausibility, during on-line sentence processing (e.g., Williams, Möbius, & Kim, 2001).

In spite of these generalizations, there is currently considerable interest in the literature in examining how individual differences among learners might modulate the extent to which reading in an L2 is qualitatively similar to or different from reading in an L1. One individual variable that has been examined extensively is level of proficiency in the L2. Studies have provided converging evidence that proficiency in the L2 often interacts with linguistic aspects of the input in producing a parsing outcome (e.g., see Frenck-Mestre, 1997; Frenck-Mestre, 2002; Kilborn, 1992; MacDonald, 2006; Hopp, 2006 for the effect of proficiency on the extent to which L2 learners utilize structural cues, and Hoshino, Dussias & Kroll, 2004 for the effect of L2 proficiency on the incorporation of semantic cues). Likewise, amount of immersion experience in the L2 has been shown to alter the processing strategies of L2 learners and counteract the influence

of L1 processing patterns when processing the L2 (e.g., Dussias & Sagarra, 2007). Another possibility, and one that is explored here, is whether cognitive capacity modulates qualitative differences and similarities between L1 and L2 sentence processing. As will be explained below, we address this question in the context of how English L1 and L2 readers with different reading span abilities are able to utilize plausibility information incrementally.

We know from the monolingual sentence processing literature that individual differences in working memory affect sentence processing behavior. Some studies have provided evidence that working memory capacity modulates the interpretation of temporarily ambiguous relative clauses (Swets, Desmet, Ferreira, & Hambrick, 2007; Kim & Christianson, 2007). Other studies have shown that individuals with high working memory capacity show a speed disadvantage when resolving certain types of syntactic ambiguity, but a speed advantage for integrating pragmatic information (Just & Carpenter, 1992; MacDonald, Just, & Carpenter, 1992; Mendelsohn & Pearlmutter, 1999). Given that the sentence processing literature provides results influenced by notions of working memory constraints, it seems important to explore whether there might also be a relationship between availability of cognitive resources and the ability, or lack thereof, to incorporate different types of information during L2 on-line sentence processing. This question becomes particularly relevant in light of previous research showing that there are increased memory demands on processing in a second language, particularly when the L2 is acquired late (e.g., Michael & Gollan, 2005; Miyake & Friedman, 1998). This might in turn constrain the ability to fully compute some semantic and syntactic nuances that are automatically incorporated during L1 processing (e.g.,

Finkbeiner, Forster, Nicol, & Nakamura, 2004; Kotz & Elston-Güttler, 2004; Silverberg & Samuel, 2004). If so, examining individual differences in cognitive resources among L2 learners will provide us with a more accurate understanding of the differences and similarities between L1 and L2 processing. Thus, we might hypothesize that the more cognitive resources L2 learners have available to them when reading in their L2, the more they might approximate sentence processing in an L1.

With this in mind, the goal of the present study is to test and contrast how native speakers and second language learners of English parse long distance *wh*-questions, what kind of information they take into consideration during real-time processing of these structures, and whether cognitive resources modulate their parsing decisions. Specifically, we focus on the extent to which working memory capacity affects the ability to utilize plausibility information in order to recover from an initial misparse in subject/object long-distance wh-structures that have been shown to produce garden path effects[1] (Juffs and Harrington, 1995).

We begin by reviewing previous research motivating the importance of processing plausibility information in order to arrive at a correct final analysis in structures that require reanalysis processes; we then report the present experiment and the findings, and the last section presents a general discussion.

II Plausibility and reanalysis

Several psycholinguistic studies in the monolingual literature have examined the role that plausibility plays in sentence processing. For example, Garnsey, Tanenhaus and

Chapman (1988) (also see Tanenhaus, Boland, Garnsey, & Carlson, 1989; Stowe, Tanenhaus, & Carlson, 1991; Tanenhaus, Stowe, & Carlson, 1985) used Event Related Potentials, or ERPs, to test whether readers would be sensitive to the plausibility of a filler with respect to a particular gap position. In their study, they contrasted the *wh*-constructions illustrated in (1)-(3):

(1) Which book did the children read __in class?

(2) *Which food did the children read __in class?[2]

(3) Which food did the children read __about in class?

Syntactic theoretical accounts posit a filler-gap dependency between the *wh*-phrase and the position directly following the verb--the canonical object position in a Subject-Verb-Object language like English. In (1) above, “which book” is a plausible object for the verb “read,” whereas “Which food” in (2) and (3) is not. If there is any psychological reality to the theoretical filler-gap dependency, we might expect that an anomaly should be detected right at the point at which the posited gap should be filled with the available *wh*-phrase. In fact, in this experiment the magnitude of the N400 effect that is typically associated with the perception of semantic anomaly was detected at the point immediately following the verb “read” in (2) and (3). The timing of the detected semantic anomaly indicated that the *wh*-phrase was interpreted at the position of the postulated gap following the verb. Additionally, the fact that the *wh*-phrase was inserted at the first permitted syntactic location (despite the fact that the meaning of the verb and that of the noun in the “which-N” expression were incompatible) indicates that plausibility neither affects the postulation of an empty category nor the establishment of a referential link to a potential antecedent.

Although semantic plausibility does not appear to affect the postulation of a gap, it has been shown to influence how fast reanalysis, when necessary, will happen. For example, Traxler and Pickering (1996), using eye-movement records, found that readers recovered faster from a garden path effect that was the result of an erroneous gap postulation when the identified filler was semantically implausible. Thus, the disambiguating region 'about while' was read faster in constructions like (4), where the potential gap-filler 'the city' is not a plausible direct object for the verb 'wrote'. In (5), where 'the book' is a semantically plausible filler for the verb 'wrote', reading times were slower at the disambiguating region.

(4) We like the city that the author wrote unceasingly with great dedication about while waiting for a contract.

(5) We like the book that the author wrote unceasingly with great dedication about while waiting for a contract.

Arguably, this effect is due to the fact that readers give up an implausible interpretation faster than a plausible one.

Recently, a number of studies have also investigated how adult L2 learners use plausibility information to recover from garden paths on-line. These studies have produced empirical evidence demonstrating the rapid influence of plausibility information during L2 sentence processing and have shown that, in this respect, non-natives can behave in a native-like way. For example, Frenck-Mestre and Pynte (1997) investigated how advanced English learners of French and French native speakers resolved prepositional-phrase attachment ambiguities in sentences such as *They accused the ambassador of espionage (of Indonesia) but nothing came of it*. Eye-movement

records revealed that both groups of speakers were influenced by plausibility information. Participants were more likely to attach the prepositional phrase to the verb phrase if it was a plausible verbal argument (e.g., *accused...of espionage*), but to the noun phrase when it was a plausible NP modifier (e.g., *ambassador of Indonesia*). Similarly, Felser and Roberts (2004) found that Greek learners of English were strongly influenced by plausibility and had difficulty recovering from misanalysis when deciding if a post-verbal NP functioned as a direct object or as an embedded subject (e.g., The man read the book (girl) had upset very many people). In particular, participants experienced processing difficulty if the initial analysis of the ambiguous NP as a direct object led to an implausible semantic interpretation.

The question of whether plausibility plays a role for L2 speakers in the processing of filler-gap structures has been addressed by Williams, Möbius and Kim (2001). They explored native-speaker and non-native-speaker differences regarding whether the semantic plausibility of a potential filler modulated the postulation of a gap during parsing. Their study, which involved native English speakers and native speakers of Korean, Chinese and German, compared the processing of sentences like (6) and (7) using a self-paced, plausibility judgment task. In this task, participants read sentences on a computer screen one word at a time and were asked to press a button as soon as they thought the sentence no longer made sense.

(6) Which girl did the man push the bike into late last night?

(7) Which river did the man push the bike into late last night?

The findings showed that both native and non-native participants were more likely to make stop-making-sense decisions at the verb site in sentences like (7), where the *wh* -

phrase ‘Which river’ was an implausible filler as the object of the verb ‘push’, thus providing evidence for gap postulation and simultaneous incorporation of semantic information. Interestingly, they also found that both native and non-native speakers showed slower reading times at the post-verbal noun-phrase region (‘the bike’) in sentences like (6), where the *wh*-filler is plausible (*Which girl did the man push?*). They argue that this is because when the filler is plausible as the direct object of the verb, it is more costly to discard it as the actual gap filler. By contrast, when the *wh*-filler is implausible as the direct object of the verb, as in (7) (*Which river did the man push?),* there is less resistance to reanalysis and, therefore, reading times are faster at the position of the real filler (‘the bike’). This comparison between native and non-native English groups suggests that adult learners of English parse *wh*-questions using strategies that are very similar to those adopted by English native speakers, even when the parallel structures in their native languages look very different.

In a subsequent study, Williams (2006) reports that incremental interpretation of plausibility information in an L2 is task dependent. In two separate experiments, participants were required to read English sentences presented one word at a time and to perform one of two tasks: (1) A stop making sense task and (2) a memory probe task that required the completion of a sentence using a word that had appeared in a previously displayed stimulus. The results showed that English-L2 learners processed plausibility information incrementally, just like the native English speakers did, when the task encouraged such type of processing (i.e., in the stop-making-sense task). However, when the task imposed memory demands, not all readers used plausibility information nor did they use it at the same point during the processing of the sentence. The native English

speakers who were the most successful in the memory probe task used plausibility information early in the sentence. Native speakers of English whose performance on the memory probe task was lower incorporated plausibility information only at a later point. Finally, only the English L2 learners with higher performance in the memory task used plausibility information, but did so later than both English native groups. Although Williams did not administer an independent measure of cognitive resources (e.g., a working memory test), the cline in the incorporation of plausibility information in the memory-demanding task suggests that the non-native readers had more difficulty processing the input incrementally most likely because they were not able to allocate sufficient cognitive resources to perform such type of processing.

It is worth noting that studies that have examined the relationship between availability of cognitive resources and sentence processing in a second language have heretofore not yielded results supporting the existence of such a relationship (e.g., Juffs, 2004 and 2005; Felser & Roberts, 2007). However, the results in Williams (2006) are suggestive enough to motivate further investigation of whether there is a relationship between individual cognitive resources and the processing of plausibility information in an L2. With this purpose, we administered an independent reading span task (i.e., a version of Waters & Caplan's 1987 and 1996 reading span test) to test whether the pre-existing cognitive characteristics of the learners influence the ability to access plausibility information with a task that does not explicitly call attention to plausibility (a grammaticality judgment task). The details of the study are presented in the next section.

III The present study

To address the questions posed above, we designed a word-by-word reading task in which the participants were asked to make grammaticality judgments on long-distance *wh*-extraction structures similar to the ones analyzed by Juffs and Harrington (1995), such as (8) and (9):

(8) Who$_i$ did the police know the pedestrian killed t_i?

(9) Who$_i$ did the police know t_i killed the pedestrian?

For the L2 group, we tested native speakers of Chinese, a language without overt *wh*-movement. Long-distance *wh*-questions are a good testing ground for two reasons. First, they have been shown to produce initial misanalysis (or a garden-path effect), given that the *wh*-filler might be temporarily interpreted as the object of the main clause (e.g., "Who did the police know?") before more material from the embedded clause becomes available. They provide, thus, an opportunity to examine how L1 and L2 readers deal with misanalysis and what kind of information they use to reanalyze the structure. A second reason why the long-distance *wh*-questions in this study are of interest for our research purposes is that they pose a challenge to the parser in that a considerable distance exists between the *wh*-phrase and the gap position where it needs to be interpreted. The individual parsing the sentence, therefore, has to keep a large amount of information in working memory in order to interpret the displaced constituent in its canonical position. For L2 speakers, this structure might pose an extra burden on the already high working memory load imposed by the task of reading in a second language. The structure, therefore, provides an ideal context in which to examine whether

individual cognitive resources provide an advantage when reading in an L2 by facilitating access to potentially useful semantic information during syntactic reanalysis.

Evidence that *wh*-extraction structures result in a garden-path comes from a well-known subject/object extraction asymmetry effect first reported in L2 learners of English by Juffs & Harrington (1995, 1996). In that study, and in subsequent replications of the results (e.g., White & Juffs, 1998), object extraction structures such as (8) were easier to process than subject extraction structures, as in (9). One explanation of this phenomenon is that the parser employs information about a verb's thematic roles to make early structural commitments during sentence processing, and constantly updates the syntactic structure it assigns to a string of words in accordance with principles of syntax (e.g. the *Generalized Theta Attachment (GTA)* of Pritchett, 1992). One prediction made by the GTA is that there will be a cost for reanalysis when it involves changes in theta and Case properties of an A-bar chain (Juffs & Harrington, 1995, 1996; but see Juffs, 2005 and Jackson & Dussias, in press, for counterevidence to some of the arguments raised in Juffs, 2005). In (8) above, for example, the noun phrases that have appeared by the time the main verb is encountered will be evaluated with respect to the possible thematic roles associated with the verb. Thus, when 'know' enters the parse, its argument structure becomes available, and the arguments 'who' and 'the police' are provisionally assigned the 'theme' (object) and 'agent' (subject) thematic roles, respectively (i.e., *Who did the police know?*). This analysis becomes untenable when the next word 'the' is processed, and the parser is forced to reanalyze the filler as the object of 'killed.' This requires a change in theta role assigner, from 'know' to 'killed,' but leaves the theme theta role of 'who' intact. By contrast, when a *wh*-word is extracted from the subject position in

sentences such as (9), reanalysis should be more costly (as evidenced by longer reading times). As in the previous case, when the argument structure of 'know' is accessed, 'who' and 'the police' are assigned the roles of 'theme' and 'agent.' However, when 'killed' is encountered, the initial analysis must be relinquished and the parser must restructure the string to allow the *wh*-filler to be the subject of a new clause. Such reanalysis requires, in addition to a change in theta/Case assigner, a change of the features of the A-bar chain: (a) a change in theta role (from internal to external) and (b) a change in case (from 'accusative' to 'nominative'). It is, therefore, presumed to be more costly than extraction from an object position, which only requires a change in theta/Case assigner. [3]

Differences in reading time between subject and object extraction sentences at the region following the main verb would indicate that a misanalysis has occurred and that the parser embarked on a process of reanalysis to arrive at the correct interpretation of the sentence. We will compare subject and object extraction structures in order to first establish that a misparse has occurred. Additionally, although, as we discussed in the previous section, plausibility has been shown to not prevent the postulation of a filler-gap dependency, we examine whether the plausibility of the postulated filler aids in the reanalysis process. To this end, we constructed sentence pairs illustrating one case in which the *wh*-phrase could be plausible as the object of the main verb and one case in which the *wh*-phrase would be an implausible filler as the object of the main verb. Based on previous findings, the prediction is that an implausible initial parse should make reanalysis faster. We will examine whether such a prediction is borne out for both monolingual English speakers and for high and low span L2 speakers.

1 Method

Participants

Eighty-four Chinese-English participants, all students at a large American university and 34 functionally monolingual English controls, also students at the same institution, were recruited. Of the 84 Chinese-English participants who completed the experiment, 23 were eliminated from the analysis after their responses to the grammatical and ungrammatical stimuli in the experiment proper were screened for YES or NO response biases. Thus, some participants might have appeared to be accurate in judging the grammatical sentences in that their response was YES for a high percentage of these items. Examination of their responses to the ungrammatical sentences, to which the correct response was NO, was critical in determining whether their YES responses to the grammatical items were due to true accuracy or to a general bias to respond YES. Five Chinese-English participants were eliminated from the analysis due to technical errors. Also, for 11 additional participants, there were no correct trials in one or more of the four conditions. Because this made reaction time analysis untenable, it was decided that these participants would be omitted from the analysis. Finally, participants who did not achieve an overall accuracy of 60% or more in the grammaticality judgment task for the *wh*-extraction constructions were excluded from the analysis (see Juffs, 2005). This eliminated 9 additional participants. Thus, our results for the Chinese-English L2 speakers are based on a total of 36 participants. All participants were paid $15 for their participation.

At the time of data collection, the 36 Chinese-English participants were completing graduate coursework either in the humanities or in the sciences. Their age

ranged from 24 years-of-age to 30 years-of-age. For these participants, acquisition of Chinese (their first language) had been completed both through formal and informal input before the onset of second language acquisition. Participants reported having received formal instruction in English beginning at age 12.

Given that for this population English is not a native language, scores from the Test of English as a Foreign Language (TOEFL) with a minimum score of 550 (computer score 213) are required by the university for admission. This ensures that their linguistic capabilities are adequate to function in academic environments. According to their TOEFL score, the 36 Chinese-English participants were proficient in English.

In addition, the participants were asked to complete a language history questionnaire, which asked questions about language dominance, self-rated level of proficiency in the four language areas (with 1 being the lowest score and 10 being the highest score), number of years the second language was studied, length of stay in a country where the second language was spoken, degree of acculturation/integration in the second language environment, and so forth.

The language history questionnaire served two functions. One was to obtain preliminary information to screen participants. Only those Chinese-English speakers with an overall self-rating proficiency score in English of 7.5 or above were called back to complete the experiment (average self-rating scores for the 36 participants whose data was included in the analysis are summarized in Table 1). The second purpose was to assess functional proficiency in the first and second languages.

------Insert Table 1 about here------

As Table 1 shows, the Chinese-English speakers considered themselves more proficient in the second language in linguistic tasks involving reading and listening than in those involving speaking or writing. The language history questionnaire also revealed that these participants had been living in an English-speaking environment for an average of 3.7 years, and normally spent an average of 6.5 hours daily reading English texts of different genres (e.g., journal articles, literary texts and newspaper articles).[4]

Finally, the 34 monolingual English participants were drawn from the student population at the same institution. None of them indicated functional proficiency in a second language above a score of 2 out of the possible 10.

2 Materials

Experimental stimuli

The experimental sentences were based on the structures used in Juffs and Harrington (1995). The majority of the verbs in the main clause were chosen from a list of 100 verbs that had been previously normed by S. Garnsey and her colleagues (personal communication) to determine verb bias information. Verb bias refers to the frequency with which a particular verb appears with a given argument. For example, *believe* can be followed by a noun phrase complement (*I don't believe <u>the boy</u>*) and by a embedded clause (*I don't believe <u>the boy arrived on time</u>*). However, it is more frequently followed by an embedded clause. There is some controversy in the literature about whether verb bias can affect the process of gap positing. For instance, Stowe et al. (1991) found that gaps are not posited after verbs that occur more frequently in intransitive frames.

Pickering and Traxler (2003) as well as Staub (2007), on the other hand, report that the parser does not initially use verb bias to posit and fill upcoming gaps. For our present purposes, it was important to ensure that our findings would not result from a combination of the plausibility manipulation in addition to verb bias effects. Hence, to minimize the impact of verb bias, we balanced our materials as much as possible with respect to the main verb's availability of a direct-object complement reading or a sentential complement reading. In all, direct-object bias verbs appeared 24 times across the experimental stimuli, sentential-complement bias verbs appeared 22 times, and equi-biased verbs—i.e., verbs that are just as likely to be followed by a direct object as by a sentential complement—appeared 13 times. No verb bias information was available for the remaining 5 verbs used in the experiment.

Once the verbs were chosen, we determined the selectional restrictions of each by asking 5 linguistically-trained judges to provide ratings regarding whether the types of noun phrase objects following each main verb either fit or did not fit the semantic constraints imposed by the verbs. Using this information, we established whether the *wh*-word heading each phrase was a plausible or implausible filler for the gap position.

Forty-three verbs were employed to write a total of 64 sentence pairs. In half of the pairs (64 sentences), the *wh*-word could be a plausible object filler for the main clause verb, whereas in the remaining half, the *wh*-word was implausible. Crossed with plausibility was extraction site. Hence, half of the 64 plausible constructions (32 sentences) contained grammatical long-distance subject extractions, and the other half contained grammatical long-distance object extractions. The same was true for the 64

implausible constructions. Examples of each type of sentence in each condition are given in (10) and (11), below. A complete list of the materials is provided in Appendix A.

The sentences in (10) exemplify the condition in which the *wh*-filler is an implausible NP object filler for the main verb. Thus, in (10a) and (10b) "Who did the police declare?" is not a plausible sentence. The sentences in (11) exemplify the condition in which the *wh*-filler is a plausible NP object for the main verb. Thus, in (11a) and (11b), 'who' could be initially parsed as the object of the main verb, as in "Who did the police know?"

Implausible filler for the main verb condition:

(10a) Who_i did the police declare t_i killed the pedestrian? (subject extraction-implausible)

(10b) Who_i did the police declare the pedestrian killed t_i? (object extraction-implausible)

Plausible filler for the main verb condition:

(11a) Who_i did the police know t_i killed the pedestrian? (subject extraction-plausible)

(11b) Who_i did the police know the pedestrian killed t_i? (object extraction-plausible)

In addition to the experimental sentences, 32 grammatical declarative sentences and 64 filler sentences, all of which were ungrammatical, were also constructed. The filler sentences contained different types of *wh*-extraction violations, such as extraction from a complex noun phrase, extraction out of a relative clause, extraction out of a sentential

subject, and extraction out of an adjunct clause. These are exemplified below in (12)-(15):

(12) * Who$_i$ does your sister believe the story that Tom married t_i? (complex NP island)

(13) *What$_i$ did the store owner see the boy who stole t_i? (relative clause island)

(14) * Who$_i$ did a story of t_i please the children? (subject island)

(15) * Who$_i$ did your wife see Tom after he met t_i? (adjunct clause island)

Four 128-item lists were constructed, so that each participant received each treatment condition in the context of a different item. In each list, there were 32 experimental sentences (8 subject extraction + plausible, 8 object extraction + plausible, 8 subject extraction + implausible, and 8 object extraction + implausible), the 32 declarative sentences and the 64 filler sentences. Each list consisted of eight blocks, and each block contained 4 experimental sentences (1 sentence representing each of the four experimental conditions), 4 declarative sentences and 8 filler sentences. Half of the sentences were grammatical and the remaining half were ungrammatical. The sentences within each block were scrambled, with the proviso that no two experimental sentences could be displayed consecutively. The order in which the blocks were presented was also scrambled.

Reading span test

To take an independent measure of their reading span, participants were asked to read eighty sentences taken from Waters, Caplan, and Hildebrandt's (1987) and Waters and

Caplan's (1996) version of the reading span task. In this test, participants are asked to read a syntactically complex sentence and to provide a plausibility judgment within a pre-determined amount of time (5000 ms in the present experiment), as well as to remember the last word of every sentence for a recall task. This test provides an index of efficiency during sentence processing, as participants need to perform a complete syntactic and semantic parse of the stimuli to provide a correct plausibility judgment. Presumably, participants with lower cognitive resources will experience difficulty performing these operations under demanding time constraints. The test also gave us a measure of the participants' working memory, since it required them to allocate cognitive resources to the temporary storage of the words for the recall task. Finally, because our main research question asked whether second language speakers are able to access and use plausibility information to recover from syntactic misanalysis, it was important to ensure that they were, in fact, sensitive to this kind of information. We were able to determine this by examining the number of correct judgments provided by the participants during the plausibility judgment task in the reading span task.

Half of the sentences in the reading span task were semantically plausible and the remaining half were semantically implausible. The 80 sentences were categorized into four types of syntactic structures:

(16) It was the neighbor that destroyed the garden. (Cleft Subject)

(17) It was the robbery that the police officer prevented. (Cleft Object)

(18) The tennis player hit the ball that was out of bounds. (Relative Subject)

(19) The package that the driver delivered thrilled the recipient. (Relative Object)

Implausible sentences were created by inverting the animacy of the subject and object nouns (e.g., *It was the police officer that the robbery prevented*). The 80 sentences were grouped into five sets (span sizes 2-6 sentences). Each span size was presented four times.

3 Procedure

Participants were tested individually in a quiet room. They were seated in front of a computer monitor. All participants were given the experiment proper first, followed by the reading span task.

Moving Window Experiment

The stimuli were displayed using a non-cumulative reading moving window technique (Just, Carpenter, & Woolley, 1982). Participants first received written instructions on the computer monitor. They were informed that for each trial, they would see a line of dashes representing each word of the sentence. Each click of the space bar would change the line of dashes into the next word and would make the previous word disappear. Each word appeared in its corresponding position within the sentence, while the position of all previous and subsequent words remained indicated on the screen by the place-holding dashes. On each trial, a fixation sign (+) was presented at the center of the computer screen. At the press of the space bar, the fixation sign disappeared and the first word of the sentence appeared left-justified. Participants were informed that their task was to read each word silently and to press the spacebar to display each consecutive word on the

screen. One example was given in the instructions. The time between the appearance of each word and the press of the space bar was recorded.

When participants reached the end of each sentence, they were asked to perform a grammaticality judgment task, by pressing a YES or a NO button accordingly. Prior to the experiment, the notion of grammaticality and ungrammaticality was explained to the participants, and examples of grammatical and ungrammatical sentences in English (involving structures other than the ones targeted in the experimental stimuli) were given to them to ensure that they understood what they were expected to judge. To familiarize participants with the task, six practice trials preceded the actual experimental session.[5]

Reading span test

Participants received written instructions on the computer screen. They were asked to read silently a series of sentences that appeared, one by one, on the screen, and to decide whether the sentences made sense or not while, at the same time, remembering the last word of each sentence. Participants initiated each trial by pressing a *yes* button (on the left) or a *no* button (on the right). A fixation sign (+) appeared immediately after the button press and remained at the center of the screen for 500 ms. The fixation sign was then replaced by a sentence. Participants were asked to read the sentence silently and to judge its plausibility. If it was semantically plausible, they pressed the “yes” button; if it was not, they pressed the “no” button. After they responded, a fixation sign again appeared for 500 ms and another sentence was presented. At intervals ranging from two to six sentences, the word *RECALL* appeared on the computer screen. Participants were asked to write down the final word of each of the previous sentences that they had just

read. The reaction time and accuracy in making judgments about semantic plausibility were recorded by the computer. Three practice blocks of 2, 3 and 4 sentences each preceded the experimental session.

Each participant response was coded for judgment accuracy and RT. Words were recorded as correct, unrecalled, or intrusions (any words other than those presented sentence-finally). Mean RTs were calculated across each participant for those items correctly assessed as plausible or implausible, and outliers 2.5 standard deviations above or below the mean were trimmed. Next, the number of correctly recalled words corresponding to correctly assessed sentences was tallied. The resulting number was used as an index of the participant's reading span.

IV Results

1 *Grammaticality judgment task*

Ungrammatical sentence judgments

The mean overall rate of correct rejection for the four types of ungrammatical *wh*-extraction constructions was calculated for the English monolinguals and the Chinese-English speakers. Accurate responses (i.e., ungrammatical sentences that were correctly rejected) were assigned a score of '1' and inaccurate responses where given a score of '0'. Results indicate that the percentage of correct responses for the English monolingual group was significantly higher (M = 96.69%) than for the Chinese-English group (M = 88.71, t (68) = 4.32, $p < 0.001$), indicating that the monolingual speakers are more accurate than the Chinese-English speakers at correctly rejecting ungrammatical *wh*-

extractions. Overall, the pattern of response accuracy for the ungrammatical items is similar to the one reported in Juffs and Harrington (1995), for stimuli that were presented using a moving window paradigm. There, the English control group correctly rejected ungrammatical *wh*-extractions at an overall rate of 93%, whereas the Chinese-English speakers correctly rejected ungrammatical sentences 82% of the time. Although in both studies the performance of the monolingual English group is superior to that of the Chinese-English speakers, the important result here is that the Chinese-English speakers are correctly rejecting ungrammatical *wh*-sentences well above chance. The high percentage of correct rejections of ungrammatical *wh*-extractions is evidence that Chinese-English speakers have knowledge of restrictions on *wh*-movement in English.

Grammatical sentence judgments

As in the previous case, a score of '1' was assigned when the sentence was judged 'grammatical' and a score of '0' when it was judged 'ungrammatical'. For declarative sentences, the average accuracy for the monolingual English speakers was 95.4% (*SD* = 4.63). The Chinese-English participants were also highly accurate in judging the declarative sentences (*M* = 93.83, *SD* = 5.99). This is an unsurprising result, given that declarative constructions that are presumed not to involve anaphoric co-reference, structural ambiguities, empty categories or other characteristics that are known to disrupt processing, are not expected to pose difficulty for readers. The average accuracy in judging grammatical *wh*-extractions for the native English speakers was 79.68% (*SD* = 9.36). For the Chinese-English participants, the average was 72.51% (*SD* = 9.83). There is a clear decline in accuracy between declarative and *wh*-extraction structures for both

the native English speakers and the Chinese speakers. For the English native group, there was a relative decline of 16% from declarative to grammatical *wh*-extraction constructions, and a relative decline of 17% from ungrammatical to grammatical *wh*-extraction. For the Chinese-English group, the decline was 22% from declarative to grammatical *wh*-extraction, and 18% from ungrammatical to grammatical *wh*-extraction. The decline in accuracy when judging grammatical *wh*-extractions suggests that these types of sentences pose parsing difficulties for both the English native and the Chinese-English L2 speakers.

Looking in more detail at the accuracy of the participants' responses to the grammatical *wh*-extraction items (Table 2), we find that for the monolingual and L2 groups, accurate judgments in sentences with plausible fillers was lower than judgments provided when the filler was implausible.

---Insert Table 2 here---

A mixed Analysis of Variance (ANOVA) with plausibility (plausible filler versus implausible filler) and extraction site (subject extraction versus object extraction) as the within-participants factor and group (English-L1 versus Chinese-English speakers) as the between-participants factor showed a significant main effect of plausibility [F (1,68) = 53.90, $p < 0.001$, MSE = 150.87] and a significant main effect of extraction site [F (1,68) = 5.92, $p < 0.05$, MSE = 157.09], but no interaction between group and any of the other two variables. One interesting finding that emerges is an asymmetry in response accuracy based on the plausibility of the *wh*-filler. In general, both groups of participants were more accurate in judging *wh*-extraction as grammatical when the filler was implausible (81.49%) than when the filler was plausible (70.7%). This result parallels the findings

reported in Stowe (1989) for monolingual English speakers. In that study, structures like '*When the police stopped the driver became very frightening'*, where plausibility of the noun phrase '*the driver*' as a direct object of the verb leads readers to incorrectly commit to this syntactic assignment, were judged ungrammatical more often than sentences such as ''*When the police stopped the silence became very frightening'*. In the second type of sentence, the implausibility of the noun phrase 'the silence' to serve as the object of the preceding verb helps readers avoid imposing an incorrect syntactic analysis on the sentence. In the present study, the decrease in incidence of incorrect responses for sentences with implausible *wh*-fillers observed for the Chinese-English speakers is compatible with the notion that for advanced second language speakers, selectional restrictions cues, such as plausibility information also have an effect on syntactic processing.

The grammaticality judgment task also revealed a difference in response accuracy based on extraction site. Participants were more accurate judging the grammaticality of sentences involving object extraction (77.92%) than subject extraction (74.27%). This result is consistent with the findings in Juffs and Harrington (1995) and suggests that misanalyses involving a noun phrase object that has to be reanalyzed as a subject cause more difficulty than a misparse involving a noun phrase object that has to be reanalyzed as the object of a lower clause.

2 Reading moving window task

Analyses were conducted on the disambiguating region, defined as the three words immediately after the main verb. To illustrate, for the subject extraction construction

(10a) *Who$_i$ did the police declare t_i killed the pedestrian?*, the embedded verb, as well as the following determiner and noun (i.e., 'killed the pedestrian') were identified as the disambiguating region. Similarly, in (10b) *Who$_i$ did the police declare the pedestrian killed t_i?*, the disambiguating region is constituted by the determiner in the embedded clause, its accompanying noun and the embedded verb (i.e., 'the pedestrian killed'). Reading times for the disambiguating region were obtained by averaging across the three words in the region. Measuring the added reaction times of the lexical material appearing after the main verb allowed us to take comparable measures between subject and object extraction structures. Comparing only the word immediately following the posited gap – 'killed' in (10a) to 'the' in (10b) – would not have made equivalent comparisons of these two types of structures feasible. This is so because it could be argued that any differences obtained between the two structures are the result of differences in the nature of the lexical items being compared (e.g., an inflected verb in (10a) vs. a determiner in (10b), rather than the result of genuine differences between the two types of structure.[6]

Following standard practice in this type of experiment, incorrect responses to the grammaticality judgment task were eliminated from the reading time analysis. In addition, prior to the analysis of the data, reading times per word below 200 ms or above 2000 ms were excluded from the analysis. No additional data trimming was performed. The same analytical procedure was used in all experiments to conduct the participants analyses (F_1). That is, a mixed ANOVA was conducted with plausibility (plausible filler versus implausible filler) and extraction site (subject extraction versus object extraction) as the within-participants factor and with group (English-L1 or Chinese-English) as the

between-participants factor. In the items analyses (F_2), extraction was a within-items variable and plausibility was as a between-items variable.

We begin by presenting the results that compare reading times in the four extraction conditions (i.e., subject extraction + plausible, object extraction + plausible, subject extraction + implausible, and object extraction + implausible) at the location immediately preceding the main verb (N-1) and at the main verb (N). Although not particularly informative for our research questions, it is important to ensure that there are no significant differences in latencies prior to the disambiguating region. Mean reading times are presented in Table 3.

------Insert Table 3 about here------

For reading times at N-1, there was no main effect of extraction, no main effect of plausibility and no interaction between group and either of the other two variables (all *F*s < 1). Similarly, reading times at N resulted in a lack of a main effect of extraction, no main effect of plausibility, and no interaction between group and either of the other two variables (all *F*s < 1). These results indicate that the material prior to the display of the critical region did not pose problems for either the monolingual participants or the Chinese-English L2 speakers. This result is expected because theoretically no reanalysis is predicted before or at the main verb.

We now turn to the data of central interest—the reading times at the disambiguating region. A mixed ANOVA with plausibility and extraction site as within-participants factors and with group as a between-participants factor showed a main effect of extraction (F_1 (1,68) = 40.90, $p < 0.001$, *MSE* = 6754.63), an interaction between extraction and group (F_1 (1,68) = 6.29, $p < 0.05$, *MSE* = 6754.63), and a three way

interaction among extraction, plausibility, and group (F_1 (1,68) = 5.07, p < 0.05, *MSE* = 8699.69). Analyses by items revealed a significant effect of extraction (F_2 (1,94) = 16.92, p < 0.01, *MSE* = 297007.30) and no interaction between extraction and plausibility (F_2 < 1). To further inspect the nature of the three way interaction among extraction, plausibility and group, we conducted separate analyses for the monolingual and the L2 groups. First, we report the results for the monolingual group.

English Monolinguals

To examine whether the availability of cognitive resources modulates the extent to which readers use plausibility information during reanalysis, the participants' reading span score was calculated using the scoring procedures described earlier, and a median split was used to group the higher working memory monolingual English participants and the lower working memory ones. A mixed ANOVA with plausibility and extraction as within-participants variables and working memory score as a between-participants variable revealed a significant main effect of extraction, with longer latencies at the critical region when the *wh*-word was extracted out of a subject position (493.03 ms) than when it was extracted out of an object position (455.05) (F_1 (1,32) = 6.66, p < 0.05, MSE= 5853.93; F_2 (1,30) = 7.87, p < 0.05, *MSE* = 20102.07), a significant main effect of plausibility, with longer latencies at the critical region for sentences containing plausible fillers (490.86 ms) than for sentences with implausible fillers (457.22 ms) (F_1 (1,32) = 15.76, p < 0.01, MSE= 3148.58; F_2 (1,30) = 7.12, p < 0.05, *MSE* = 143138.43). There was also a marginally significant interaction between extraction, plausibility and working memory (F_1 (1,32) = 3.97, p = 0.055, MSE= 2903.27).

Although the interaction between extraction, plausibility and working memory was only marginally significant and conclusions in this regard can only be suggestive, we report the results for the two span groups in Table 4 for illustration purposes.

------Insert Table 4 about here------

As Table 4 shows, reading time patterns for the higher working memory speakers reveal longer reading times for the three-word critical region for subject-extraction structures vs. object-extraction structures in both plausibility conditions. In addition, we observe that both in the subject and in the object extraction conditions, sentences with implausible *wh*-fillers produced lower reading times, although the significance of this difference cannot be ascertained. Turning now to the lower working memory speakers, they also appear to have the most difficulty reading subject-extraction structures with plausible *wh*-fillers. However, in contrast to their higher reading span counterparts, they display relatively equal difficulty processing the three other conditions. These results hint in the direction that only the higher span natives are able to use plausibility information in both contexts.

Next, we examined at what point in the sentence the plausibility of the *wh*-word as a potential filler for the gap was used by readers. Word-by-word comparisons are relevant here because previous studies (e.g., Williams et al. 2001) have shown that monolingual and L2 readers differ in how early they use plausibility information to recover from misanalysis. To determine when readers used plausibility information to recover from misanalysis, we compared word-by word reading times for parallel structures; namely, subject-extraction sentences with plausible and implausible fillers (*Who$_i$ did the police declare t_i killed the pedestrian?* and *Who$_i$ did the police know t_i killed the pedestrian?*) and object extraction sentences with plausible and implausible

fillers (*Who$_i$ did the police declare the pedestrian killed t_i?* and *Who$_i$ did the police know the pedestrian killed t_i?*). The word-by-word comparison is now possible because, crucially, plausible and implausible subject-extraction structures, on the one hand, and plausible and implausible object extraction sentences, on the other hand, were identical with respect to the position of the lexical items. Because participants are presumed to initiate re-analysis at some point after the main verb, we compared reading times for each of the words in the critical three-word post-verbal region. We begin by presenting the results of the subject-extraction sentences. As Figure 1 shows, there is a peak in reading times at the first position following the main verb (e.g., at "killed" in *Who$_i$ did the police declare/know t_i killed the pedestrian?*) for both plausible and implausible conditions, although reading times are longer in the plausible condition (514.27 ms.) as compared to the implausible one (464.55 ms). This difference was statistically significant (t (33) = 2.17, $p < 0.05$), suggesting that, as shown in previous studies, a plausible reading is harder to give up than an implausible reading. At the second position (i.e., "the"), there is a decrease in reading times for the plausible (445.45 ms) and the implausible (412.42 ms) conditions. The difference between the two means, however, continues to be statistically significant at this position (t (33) = 2.15, $p < 0.05$), suggesting that the effect of plausibility in the reanalysis process spills over from the first word to the second word. Finally, at the third word, reading times increase in both conditions, but reading times between the plausible (594.96 ms) and implausible conditions (527.94 ms) are not significantly different from each other (t (33) = 1.66, $p > 0.05$), suggesting that reanalysis is completed by this point. The increase in reading times, then, is most likely due to the well-documented end-of-sentence effect.

------insert Figure 1 about here------

Turning to the object-extraction structures, reading times for the plausible and implausible conditions are virtually identical to each other at each of the three words following the main verb (all Fs < 1). Unlike subject-extraction structures, the plausibility of the wh-word as a potential filler for the gap does not appear to influence reanalysis (see Figure 2).

------insert Figure 2 about here------

Chinese-English L2 speakers

As we did for the English-native group, a median split was used to group the higher working memory Chinese-English participants and the lower working memory ones. A mixed ANOVA with plausibility and extraction as within-participants variables and with reading span as the between-participants variable revealed a main effect of extraction (F_1 (1,34) = 26.17, $p < 0.001$, MSE = 10530.21), an interaction between plausibility and extraction (F_1 (1,34) = 4.20, $p < 0.05$, MSE = 9993.04), an interaction between plausibility and reading span score (F_1 (1,34) = 15.77, $p < 0.001$, MSE = 17700.72), and an interaction among plausibility, extraction, and reading span score (F_1 (1,34) = 14.74, $p < 0.01$, MSE = 9993.04). Items analyses revealed a significant effect of extraction (F_2 (1,62) = 14.44, $p < 0.01$, MSE = 420453.05) and no interaction between extraction and plausibility ($F_2 < 1$). To better understand the nature of the interaction obtained in the participants analyses, we conducted separate analyses for the L2 speakers with higher and lower reading span scores.

Reading latencies for the two groups of L2 speakers are given in Table 5. L2 speakers with higher reading span scores showed longer reading latencies at the critical region in subject-extraction sentences (M = 757.16 ms) than in object extraction sentences (M = 669 ms), (F_1 (1,17) = 18.64, $p < 0.001$, MSE = 7504.64; F_2 (1,30) = 5.23, $p < 0.05$, MSE = 481046.47). They also took longer to read the disambiguating region when the *wh*-word could be interpreted as a plausible filler for the object position of the main verb (M = 758.25 ms) than when the *wh*-word was an implausible object for the main verb (M = 667.91 ms) (F_1 (1,17) = 8.49, $p = .01$, MSE = 17289.10; F_2 (1,30) = 4.40, $p < 0.05$, MSE = 213375.2). However, there was no interaction between plausibility and extraction [all $Fs < 1$].

------Insert Table 5 about here------

Interestingly, this group's performance displays similarities to the reading latency patterns of the English monolingual participants in that both groups had longer reading times for subject-extraction structures vs. object-extraction structures in both plausibility conditions. Similarly, for both groups, sentences with implausible *wh*-fillers produced lower reading times in the subject and in the object extraction conditions.

To examine how the plausibility of the *wh*-word affected reanalysis, word-by word analyses of the post-verbal region were conducted in the manner described for the English monolingual group. Figure 3 shows that for subject-extraction structures, there is a peak in reading times at the first position after the main verb in the plausible and implausible conditions, with plausible *wh*-fillers taking significantly more time to read (894.61 ms) than implausible *wh*-fillers (744.21 ms) (t (17) = 2.02, $p < 0.05$). At the next position, reading times decrease in both conditions (677 ms for plausible fillers vs.

598.67 ms for implausible fillers); the sentences in the plausible condition display significantly longer reading times (t (17) = 2.21, $p < 0.05$). Finally, there is an increase in reading times at the last word (880.33 ms for plausible fillers vs. 748.33 ms for implausible fillers), but the difference between the two means is not statistically significant ((t (17) = 1.58, $p > 0.05$).

------insert Figure 3 about here------

Turning to the object-extraction structures (Figure 4), there are no differences in reading times for the plausible and implausible conditions at each of the three words following the main verb (all *Fs* <1). Unlike the effect found in subject-extraction structures, the plausibility of the *wh*-word as a potential filler for the gap does not seem to influence reanalysis.

------insert Figure 4 about here------

The overall pattern of results thus far indicates that the reading patterns of the higher-span Chinese-English participants are similar to those of the English monolingual group, the only difference being that the L2 participants display overall longer reading times than the monolingual speakers.

Results on the same dimensions for the Chinese-English speakers with lower reading span scores (Table 5) present quite a different picture. The data yielded a main effect of extraction significant by participants (F_1 (1,17) = 10.01, $p < 0.01$, *MSE* = 13555.78) and by items (F_2 (1,30) = 16.70, $p < 0.05$, *MSE* = 214364.21), a main effect of plausibility that was significant only by participants (F_1 (1,17) = 7.31, $p < 0.05$, *MSE* = 18112.34) and an interaction between plausibility and extraction significant by participants (F_1 (1,17) = 18.95, $p < 0.001$, *MSE* = 9149.90) and by items (F_2 (1,30) =, $p <$

0.05, *MSE* = 265362.65). Unlike their higher span Chinese-English bilingual counterparts, these speakers only showed a contrast in reading times between the subject and the object extraction conditions when the filler was implausible. Mean reading time at the critical region was 915.14 ms in the subject extraction, implausible filler condition versus 730.16 ms in the object extraction, implausible filler condition. In contrast, reading latencies in the plausible filler conditions were markedly similar to each other (731.20 ms for subject extraction and 742.52 ms for object extraction) and to the object extraction structures in the implausible condition. This suggests that the statistical effects observed for this group are a result of the higher reading times for subject-extraction sentences in the implausible condition.

Word-by-word reading times for subject-extraction structures (Figure 5) show a peak in reading times at the first position after the main verb in the plausible and implausible conditions, with implausible *wh*-filler constructions taking significantly more time to read (1099.5 ms) than plausible *wh*-filler constructions (792.72 ms) (t (17) = 3.96, $p < 0.01$). This is exactly the opposite pattern from that discussed for the two other groups of participants. At the next position, reading times decrease in both conditions (614.06 ms for plausible fillers vs. 698.89 ms for implausible fillers), but the mean difference in the two conditions is not statistically reliable(t (17) = 1.10, $p > 0.05$). Finally, there is an increase in reading times at the last word (787.22 ms for plausible fillers vs. 958.06 ms for implausible fillers); the difference between the two means is statistically significant (t (17) = 2.26, $p < 0.05$).

------insert Figure 5 about here------

Finally, for object-extraction structures (Figure 6), there were no differences in reading times for the plausible and implausible conditions at each of the three word positions following the main verb (all *Fs* <1).

------insert Figure 6 about here------

The finding that the lower span Chinese-English group does not display a subject-object extraction contrast across both plausibility conditions (as do the monolingual group and the high span Chinese-English group) indicates that these participants do not integrate syntactic and semantic information in the same way. At this point, we can only speculate as to whether or how the lower span English-L2 group utilized the available syntactic and semantic cues to process the target structures. The fact that the subject/object contrast is not consistently found across conditions might suggest that their syntactic processing is shallow and, therefore, they do not pursue a filler-gap strategy. On the other hand, they do seem to use plausibility information to some extent, given the markedly elevated reading times that they display in the subject-implausible condition. The expectation, based on previous experimental results, that implausible fillers should facilitate processing is dependent on the assumption that the processor follows a specific syntactic strategy that leads to the reanalysis of a filler-gap dependency. If the lower-span readers do not follow such a strategy, the seemingly counter-intuitive finding that implausible fillers disrupt processing is not altogether surprising.[7] Further testing is needed to arrive at a better understanding of how lower-span readers might process structures that pose a memory strain, such as the long-distance gap structures examined here. What is clear is that the lower-span L2 learners do not integrate information in the

same way or for the same purposes as the native-English speakers or their higher-span Chinese-English counterparts.

IV Discussion

One of the standing debates in the L2 literature concerns what kind of information comes into play when processing sentences in an L1 and an L2 and whether individual differences might affect processing. In this study, we examined one type of information that might be used by L1 and L2 readers during reanalysis processes in English interrogative structures that are known to cause a misparse. Specifically, we investigated whether the use of plausibility cues, a type of lexical-semantic information that has been found to be utilized by L1 and L2 readers alike, is modulated by the availability of cognitive resources in the L2. To this end, we operationalized the cognitive resources of our English-L1 and English-L2 participants in terms of reading span, and we varied the plausibility of the *wh*-word as a potential gap filler in our target *wh*-subject and *wh*-object extraction structures. Our main finding indicates that only the high span second language learners of English in our study behave in a manner that is similar to that of the English monolingual participants in that, for both groups, subject-extraction sentences posed more difficulties than object-extraction sentences in both plausibility conditions. We see this in the longer reading times for the three-word critical region for subject-extraction structures vs. object-extraction structures. Additionally, the word-by-word reading times comparing plausibility within each extraction site condition show a peak in subject extraction sentences in both plausibility conditions right at the word immediately

following the main verb, while no such peak is found in object extraction sentences. This pattern is consistent with the hypothesis that a reanalysis process takes place right at the disambiguating word (the word following the main verb), with subject extraction configurations being harder to reanalyze than object extraction configurations.[8]

Crucial to our research question, we found that in the sentences that are harder to process, namely those in which a subject was extracted, both the English monolingual group and the high-span English-L2 group displayed longer latencies when the *wh*-word was a plausible filler for the (incorrect) gap position, indicating that the participants were using plausibility information to recover from their initial misparse and that discarding a plausible reading was harder than discarding an implausible one. Interestingly, this effect was found in a task that simply asked participants to judge the grammaticality status of the sentences and did not call attention to plausibility, as do other tasks such as stop-making-sense tasks (Williams, Möbius, & Kim, 2001) or semantic-anomaly detection tasks (Tanenhaus, Garnsey and Boland, 1990). The same attention to plausibility was not found in *wh*-object extraction questions. This results may be explained if one hypothesizes that because object extraction constructions pose fewer difficulties, readers may not need to rely on other sources of information to arrive at the correct parse for the sentence. This explanation seems unlikely, however, given that other studies have shown robust filled gap effects with object gaps (Williams et al, 2001; Stowe, 1986). A second explanation is that the availability of a sentential complement reading in some of verbs used in the experiment precluded the need to use plausibility information to recover from misanalysis in *wh*-object extraction constructions. Given that the same verbs were used to construct the subject-extraction and the object-extraction pairs, and that we found a

plausibility effect in the subject-extraction constructions, we do not think that a verb bias explanation is relevant here. Additionally, as was discussed earlier, the experimental sentences were constructed using a list of verbs that had been previously normed for verb biases, and care was taken to balance the items. A third explanation for the lack of a detectable plausibility effect in object-extraction structures may have to do with syntactic priming effects that could have resulted from the repetition of some verbs in the experimental stimuli. Thirteen of the 43 verbs used to construct the experimental materials were used more than once. Four direct-object-bias verbs, three sentential-complement-bias verbs and three equi-bias verbs were each used twice. In addition, one sentential-complement-bias verb and one equi-bias verb were each used three times.[9] It is possible that the repetition of sentential-complement-bias verbs many have resulted in syntactic priming effects that in turn could have alleviated the difficulty posed by object extraction structures.[10] Although this possibility merits future investigation, what is crucial in our results is that the higher-span English-L2 participants showed the same reading pattern as the English monolingual participants (in particular to the high-span English monolingual participants) and that both groups attended to the same kind of information and employed it in the same way.

The lower-span L2 group, on the other hand, did not seem to attend to the same kind of information as the monolingual group or the higher-span English-L2 group. Although for these participants the three-word analysis shows that subject-extraction constructions appear to pose more difficulty than object extraction sentences, this effect is driven primarily by the elevated reading times in the subject extraction condition when the filler was implausible. The word-by-word measurements comparing reaction times in

the plausible and implausible conditions reveal that implausible fillers cause more difficulty than plausible fillers in the subject condition. This is the opposite from the pattern found in the monolingual English speakers and the higher-span English-L2 group, and it indicates that only at the high-span end of the spectrum do L2 speakers approximate native-like behavior, even when considering lexical-semantic information, which is presumed to be used by L1 and L2 readers alike.

To conclude, taken together, these results suggest that adequate cognitive resources are required to access and integrate different sources of information during sentence processing in an L2. While working memory has already been established as a variable that impacts language processing in an L1 (e.g., Just & Carpenter, 1992; Traxler, Williams, Blozis, & Morris, 2005), more work is needed to ascertain which factors might impact allocation of resources while processing an L2. This study indicates that individual differences in reading span among L2 learners might be an important factor. As was reviewed earlier, a recent study by Williams (2006) found that L2 speakers can process the input incrementally, just like native speakers do, when the task encourages such type of processing (i.e., in a stop-making-sense task). However, when the task imposes memory demands, non-native readers may not be able to process the input incrementally, most likely because they are not able to allocate sufficient resources to perform such type of processing. Based on an independent reading span measure, our study indicates that second language learners demonstrating high reading span capacity can, in fact, process plausibility information incrementally in a way resembling L1-language processing regardless of the task.

To summarize, this study indicates that individual differences in processing resources play a role during L2 sentence comprehension and that L2 sentence processing can be remarkably similar to L1 sentence processing provided that sufficient cognitive resources are available to perform any given task. We have established that L2 speakers with higher reading span behave more similarly to monolingual speakers in general and have pointed out the need to further explore the relationship between cognitive resources in the target language as an important individual variable in L2 sentence processing.

Acknowledgements

The research reported in this paper was supported in part by NIH Grant HD50629-1 to the first author. We thank Sarah Reed, Tracy Cramer, and Maya Misra for invaluable research and statistical assistance. Preliminary results were presented at the Second Language Research Forum 2003, the 4th International Symposium on Bilingualism, and the XIV European Society for Cognitive Psychology.

VI References

Clashen, H. and Felser, C: 2006. Grammatical processing in language speakers. Applied Psycholinguistics 27, 3-42.

Dussias, P. E. and Sagarra, N: 2007. The effect of exposure on syntactic parsing in Spanish-English L2 speakers. Bilingualism: Language and Cognition 10, 101-16.

Felser, C. and Roberts, L: 2004. Plausibility and recovery from garden paths in second language sentence processing. Poster presented at the 10th annual conference on architectures and mechanisms of language processing (AMLaP), Aix En-Provence, France.

Felser, C. and Roberts, L: 2007. Processing *wh*-dependencies in a second language: A cross-modal priming study. Second Language Research 23, 9-36.

Finkbeiner, M., Forster, K., Nicol, J., and Nakamura, K: 2004. The role of polysemy in masked semantic and translation priming. Journal of Memory and Language 51, 1-22.

Frenck-Mestre, C: 1997. Examining second language reading: An on-line look. In Sorace A, Heycock C. and Shillcock R., editors, Proceedings of the GALA ’97 Conference on Language Acquisition, Edinburgh: University of Edinburgh Press, 444–449.

Frenck-Mestre, C: 2002. An on-line look at sentence processing in the second language. In Heredia R. and Altarriba J., editors, Bilingual sentence processing, New York: Elsevier, 217-36

Frenck-Mestre, C. and Pynte, J: 1997. Syntactic ambiguity resolution while reading in second and native languages. The Quarterly Journal of Experimental Psychology 50A, 119-48.

Garnsey, S. M., Tanenhaus, M. K., and Chapman, R. M: 1988. Evoked potentials and the study of sentence comprehension. Journal of Psycholinguistic Research: Special Issues on Sentence Processing 18, 51-60.

Hoshino, N., Dussias, P. E., and Kroll, J. K: 2004. Production of subject-verb agreement in second language learners and proficient L2 speakers: Do individual differences in reading span and verbal fluency modulate their performance? Paper presented at the 4th International Symposium on Bilingualism, Tempe, Arizona, U.S.A

Hopp, H: 2006. Syntactic features and reanalysis in near-native processing. Second Language Research 22, 369–97.

Jackson, C. N. and Dussias, P. E: in press. Cross-linguistic differences and their impact on L2 sentence processing. Bilingualism, Language and Cognition.

Juffs, A: 2005. The influence of the first language on the processing of wh-movement in English as a second language. Second Language Research 21, 2, 121-51.

Juffs, A. and Harrington, M: 1996. Garden path sentences and error data in second language sentence processing. Language Learning 46, 283-326.

Juffs, A. and Harrington, M: 1995. Parsing effects in second language sentence processing: Subject and object asymmetries in wh-extractions. Studies in Second Language Acquisition 17, 483-516.

Just, M. A. and Carpenter, P: 1992. A capacity theory of comprehension: individual differences in working memory. Psychological Review 99, 122-49.

Just, M. A., Carpenter, P. A., and Woolley, J. D: 1982. Paradigms and processes in reading comprehension. Journal of Experimental Psychology: General 111, 228-38.

Kilborn, K: 1992. On-line integration of grammatical information in a second language. In Harris, R., editor, Cognitive Processing in Bilinguals. Amsterdam: Elsevier, North-Holland, 337-350.

Kim, J. and Christianson, K: 2007. Working memory effects on sentence processing. Paper presented at the Second Language Research Forum, University of Illinois, Urbana, Illinois.

Kotz, S. A. and Elston-Güttler, K: 2004. The role of proficiency on processing categorical and associative information in the L2 as revealed by reaction times and event-related potentials. Journal of Neurolinguistics 17, 215-35.

MacDonald, M. C., Just, M. A., and Carpenter, P. A: 1992. Working memory constraints on the processing of syntactic ambiguity. Cognitive Psychology, 24, 56-98.

McDonald, J: 2006. Beyond the critical period: Processing-based explanations for poor grammaticality judgement performance by late second language learners. Journal of Memory and Language 55, 381-401.

Mendelsohn A. and Pearlmutter, N., J: 1999. Individual differences in relative clause attachment ambiguity. Poster presented at the 12th Annual CUNY Conference on Human Language Development. New York, NY.

Michael, E. and Gollan, T.H: 2005. Being and becoming bilingual: Individual differences and consequences for language production. In Kroll J. F. and de

Groot, A.M.B., editors, Handbook of bilingualism: psycholinguistic approaches, New York: Oxford University Press, 389-410.

Miyake, A. and Friedman, N. F: 1998. Individual differences in second language proficiency: working memory as language aptitude. In Healy, A. F. and Bourne, L. E., editors, Foreign language learning: psycholinguistic studies on training and retention, Mahwah, NJ: Erlbaum, 339-64.

Pickering, M. J. and Traxler, M. J: 2003. Evidence against the use of subcategorization frequency in the processing of unbounded dependencies, Language and Cognitive Processes 18, 469-503.

Pritchett, B. L: 1992. Grammatical Competence and Parsing Performance. Chicago: The University of Chicago Press.

Silverberg, S. and Samuel, A: 2004. The effect of age of second language acquisition on the representation and processing of second language words. Journal of Memory and Language 51, 381-98.

Staub, A: 2007. The Parser Doesn't Ignore Intransitivity, After All. Journal of Experimental Psychology: Learning, Memory, and Cognition 33, 550-560.

Stowe, L. A: 1986. Evidence for on-line gap location. Language and Cognitive Processes, 1, 227-245

Stowe, L. A., Tanenhaus, M. K., and Carlson, G. N: 1991: Filling gaps on-line: Use of lexical and semantic information in sentence processing. Language and Speech 34, 319-340.

Stowe, L. A: 1989. Thematic Structures and Sentence Comprehension. In Carlson G. N. and Tanenhaus, M. K., (editors), Linguistic Structure in Language Processing, Dordrecht, The Netherlands: Kluwer Academic Press, 319-537.

Stowe, L. A., Tanenhaus, M. K., and Carlson, G. N: 1991. Filling gaps on-line: use of lexical and semantic information in sentence processing, Language and Speech 34, 319-40.

Swets, B., Desmet, T., Ferreira, F., and Hambrick: 2007. The role of working memory in syntactic ambiguity resolution: A psychonometric approach. Journal of Experimental Psychology: General, 136, 64-81.

Tanenhaus, M. K. , Boland, J. E., Garnsey, S. M., and Carlson, G. N: 1989. Lexical structure in parsing long-distance dependencies. Journal of Psycholinguistic Research 18, 37-50.

Tanenhaus, M. K., Garnsey, S. M., and Boland, J. E: 1990. Combinatory lexical information and language comprehension. In Altmann, G. T. A. , editor, Cognitive models of speech processing: psycholinguistic and computational perspectives. Cambridge, MA: MIT Press.

Tanenhaus, M. K., Stowe, L. A., and Carlson, G. N: 1985. The interaction of lexical expectation and pragmatics in parsing filler-gap constructions. In Proceedings of the Seventh Annual Cognitive Science Meetings.

Traxler, M. J. and Pickering, M. J: 1996. Plausibility and the processing of unbounded dependencies: An eye-tracking study. Journal of Memory and Language 35, 454-75.

Traxler, M., Williams, R., Blozis, S., and Morris, R: 2005. Working memory, animacy, and verb class in the processing of relative clauses. Journal of Memory and Language 53, 204-24.

Waters, G. S. and Caplan, D: 1996. The measurement of verbal working memory capacity and its relation to reading comprehension. Quarterly Journal of Experimental Psychology: Human Experimental Psychology 49, 51-79.

Waters, G. S., Caplan, D., and Hildebrandt, N: 1987. Working memory and written sentence comprehension. In Coltheart, M., editor, Attention and performance 12: the psychology of reading, Hillsdale, NJ: Lawrence Erlbaum Associates, Inc, 531-55.

Williams, J. N: 2006. Incremental interpretation in second language sentence processing. Bilingualism: Language and Cognition 9, 71-88.

Williams, J. N., Möbius, P., and Kim, C: 2001. Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. Applied Psycholinguistics 22, 509-40.

White, L. and Juffs, A: 1998. Constraints on wh-movement in two different contexts of nonnative language acquisition: competence and processing. In Flynn, S. Martohardjono G., and O'Neil, W., editors, The generative study of second language acquisition, Mahwah, NJ: Lawrence Erlbaum, 111-29.

Notes

[1] Descriptively, a garden path effect occurs when the parser constructs an initial incorrect syntactic analysis for an incoming string of words. When the misanalysis is recognized by the parser (due to a breakdown in the ongoing analysis), reanalysis is required. In some instances, the effort required to reanalyze is so great that readers (or listeners) are unable to reach a grammatical parse.

[2] Following the convention established in linguistics, we use asterisks to denote ungrammatical sentences.

[3] As suggested by an anonymous reviewer, an alternative explanation for the subject/object asymmetry described here is that in the object-extraction condition ('who did the police know the pedestrian killed?') 'the' disambiguates the structure, triggering disattachment of 'who' as the object of 'know.' Subsequently, 'killed' (which appears two words later) would trigger reattachment. In contrast, in the subject-extraction condition ('who did the police know killed the pedestrian?'), 'killed' triggers both disattachment of 'who' as the object of 'know' and reattachment as subject of 'killed' at the same point in time. It is argued that timing makes this second process harder. Without an explicit theory motivating why the second option would be more burdensome for the processor, we cannot commit to this analysis. However, we acknowledge that it is possible that a combination of timing constraints and the additional changes in the features of the A-bar chain that we discuss might together contribute to the asymmetry effect. What is important for our purposes is that both explanations are consistent with the notion that a misparse followed by a reanalysis process is involved in these structures.

[4] One possible way to derive an index of L2 proficiency is to correlate the participants' self-assessment measures with their TOEFL scores. However, correlation analyses were not appropriate because the participants' TOEFL scores reflected the level of proficiency they were at a number of years prior to their participation in this study.

[5] A anonymous reviewer raises the question of whether the grammaticality judgment task might have been interpreted as an acceptability judgment task, which might in turn have triggered the consideration of plausibility constraints by the participants. The researchers made sure that the notion of grammaticality was clear to the participants. In fact, the participants were highly accurate in rejecting the ungrammatical filler sentences, even though some of them were clearly interpretable and plausible, such as, for example, the *wh*-island violations (e.g., * *Who does your sister believe the story that Tom married*; * *What did the store owner see the boy who stole*). This indicates that the notions of grammaticality/ungrammaticality were understood by the participants.

[6] Defining the region of interest to include the embedded verb, a determiner, and an noun could be somewhat problematic because the lack of lexical material at the end of the sentence may produce the undesired effect of conflating any RTs effects that are reported with sentence wrap-up effects. Although this is a limitation that needs to be addressed in future research, Mitchell (2004) finds that self-paced reading studies in which the critical region is the last phrase in the sentence have produced very similar results to those obtained in eye-tracking experiments in which the stimuli have been "padded" to avoid having the region of interest coincide with the last segment in the sentence. A comparison of the findings in Dussias (2003) and Dussias & Sagarra (2007) leads to the same

conclusion. In addition, Juffs (2005) found subject-object processing asymmetries even when object extraction sentences contained lexical material after the gap.

[7] We thank an anonymous reviewer for this suggestion.

[8] This subject/object asymmetry effect replicates the overall findings in Juffs and Harrington (1995), although they did not test for plausibility or separated participants according to reading span.

[9] The thirteenth verb, for which we do not have verb bias information, was *say* and was used twice.

[10] We thank an anonymous reviewer for bringing this point to our attention.

Table 1

Language history questionnaire. Chinese-English speakers; 10 = highly proficient, 1 = minimal ability

Language area	Chinese (L1)	English (L2)
Reading	9.6	8.7
Listening	9.9	8.8
Speaking	9.7	7.9
Writing	9.4	8.0

Table 2

Percent of correct responses for grammatical and ungrammatical wh-extractions; English monolingual speakers and Chinese-English L2 speakers

	Group Type	
Sentence Type	English Monolinguals	Chinese-English L2 speakers
Ungrammatical		
M	96.69%	88.71%
SD	5.85	9.13
Subject Extraction—Plausible filler		
M	73.52 %	62.50 %
SD	16.21	15.52
Object Extraction—Plausible filler		
M	73.89 %	72.91 %
SD	20.72	11.37
Subject Extraction—Implausible filler		
M	84.55 %	76.51 %
SD	13.78	16.47
Object Extraction—Implausible filler		
M	86.76 %	78.12 %
SD	11.07	13.81

Table 3

Mean reading times in milliseconds at N-1 and N. Monolingual English speakers and Chinese-English L2 speakers

	Group Type			
Sentence Type	Monolingual English		Chinese-English L2 speakers	
	N-1	N	N-1	N
Subject Extraction—Plausible filler				
M	401.94	419.36	617.31	698.06
SD	98.19	98.60	238.63	266.51
Object Extraction—Plausible filler				
M	408.46	417.89	577.11	560.56
SD	106.83	101.44	218.70	237.18
Subject Extraction—Implausible filler				
M	414.24	420.49	610.36	612.50
SD	101.11	103.42	173.47	164.80
Object Extraction—Implausible filler				
M	402.50	408.75	577.97	578.89
SD	103.50	93.84	159.84	143.96

Table 4

Mean reading times per span score at the disambiguating region. Monolingual English Speakers

	Span	
Sentence type	Higher	Lower
Subject Extraction—Plausible filler		
M	514.54	521.01
SD	204	121.41
Object Extraction—Plausible filler		
M	482.59	445.32
SD	184.58	107.71
Subject Extraction—Implausible filler		
M	473.72	462.87
SD	152.93	94.44
Object Extraction—Implausible filler		
M	437.06	455.25
SD	134.38	82.76

Table 5

Mean reading times per span score at the disambiguating region. Chinese-English Speakers

	Span	
Sentence type	Higher	Lower
Subject Extraction—Plausible filler		
M	817.24	731.20
SD	212.50	206.59
Object Extraction—Plausible filler		
M	699.26	742.52
SD	196.96	227.51
Subject Extraction—Implausible filler		
M	697.09	915.14
SD	174.64	158.75
Object Extraction—Implausible filler		
M	638.74	730.16
SD	140.48	154.66

Figure Captions

Figure 1 Word-by-word reading times for subject-extraction structures; plausible and implausible conditions. English monolingual speakers

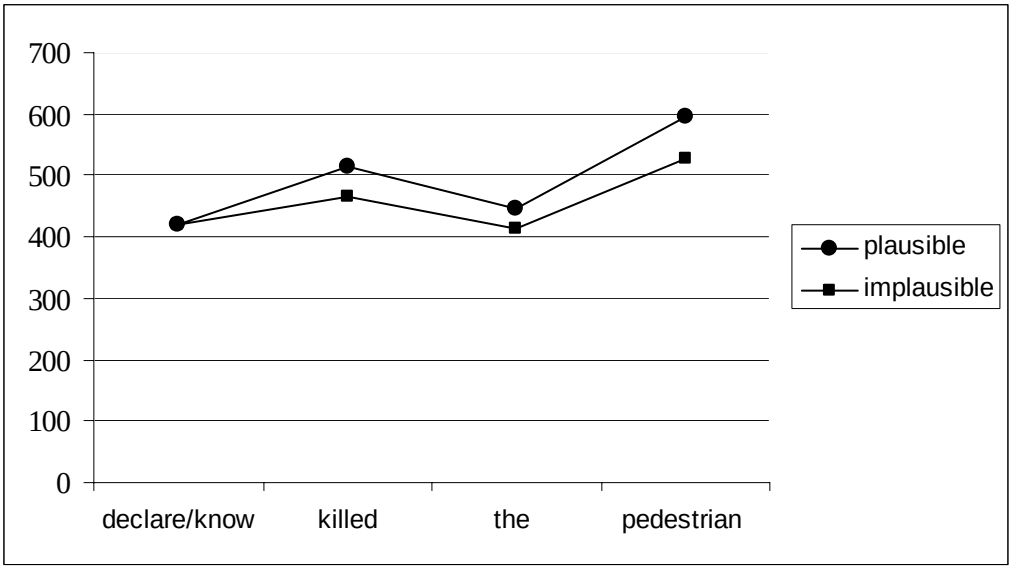
Figure 2 Word-by-word reading times for object-extraction structures; plausible and implausible conditions. English monolingual speakers

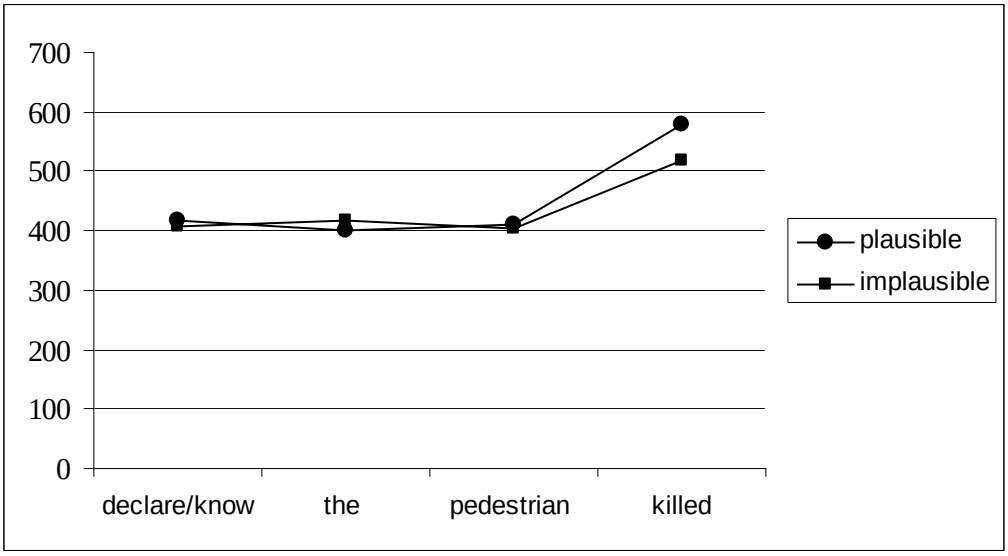
Figure 3 Word-by-word reading times for subject-extraction structures; plausible and implausible conditions. Higher span Chinese-English L2 speakers

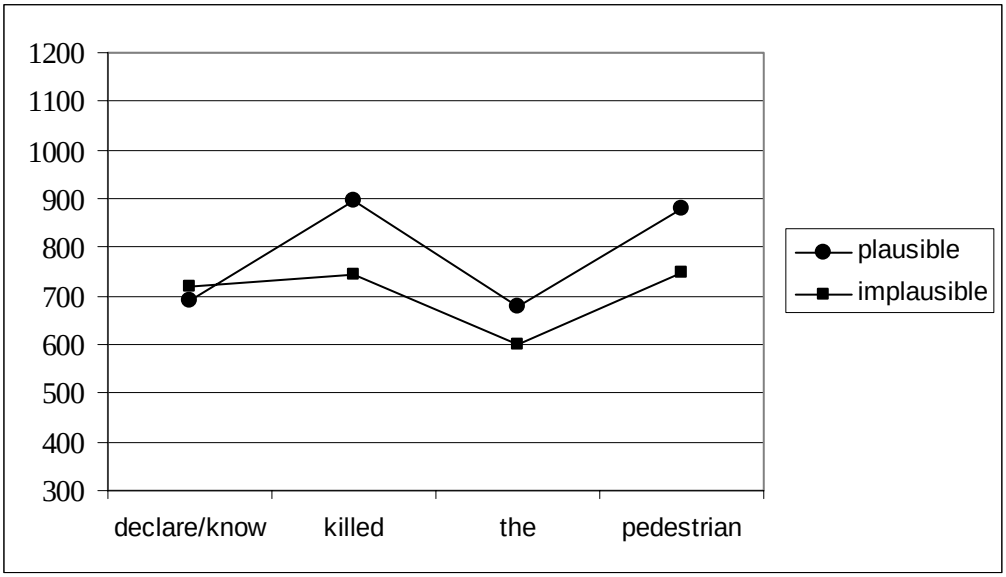
Figure 4 Word-by-word reading times for object-extraction structures; plausible and implausible conditions. Higher span Chinese-English L2 speakers

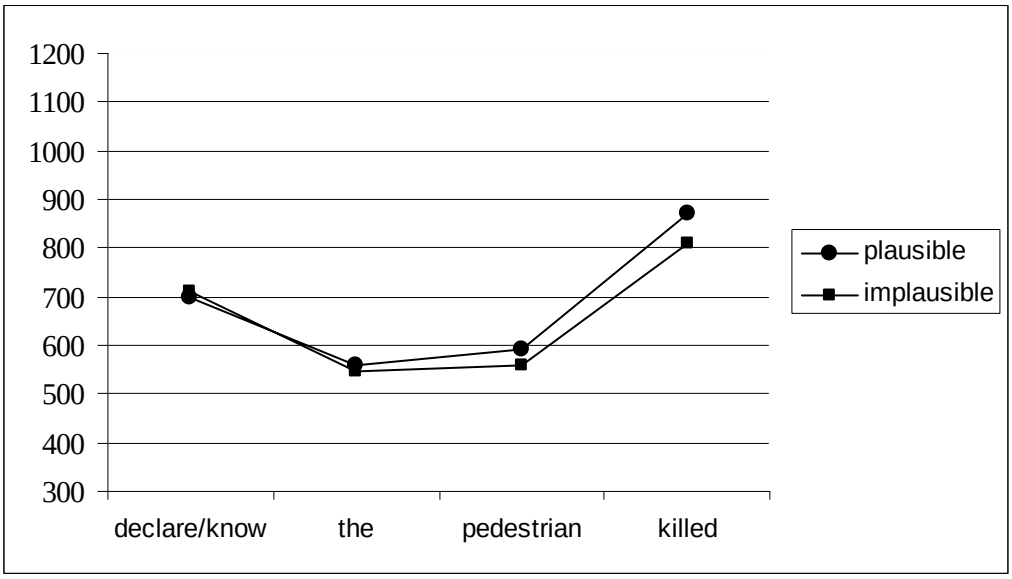
Figure 5 Word-by-word reading times for subject-extraction structures; plausible and implausible conditions. Lower span Chinese-English L2 speakers

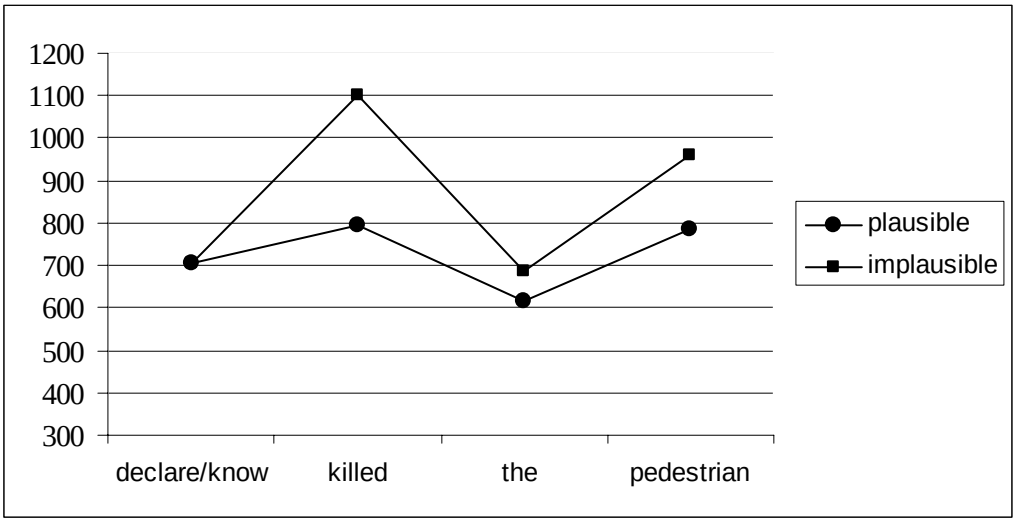
Figure 6 Word-by-word reading times for object-extraction structures; plausible and implausible conditions. Lower span Chinese-English L2 speakers

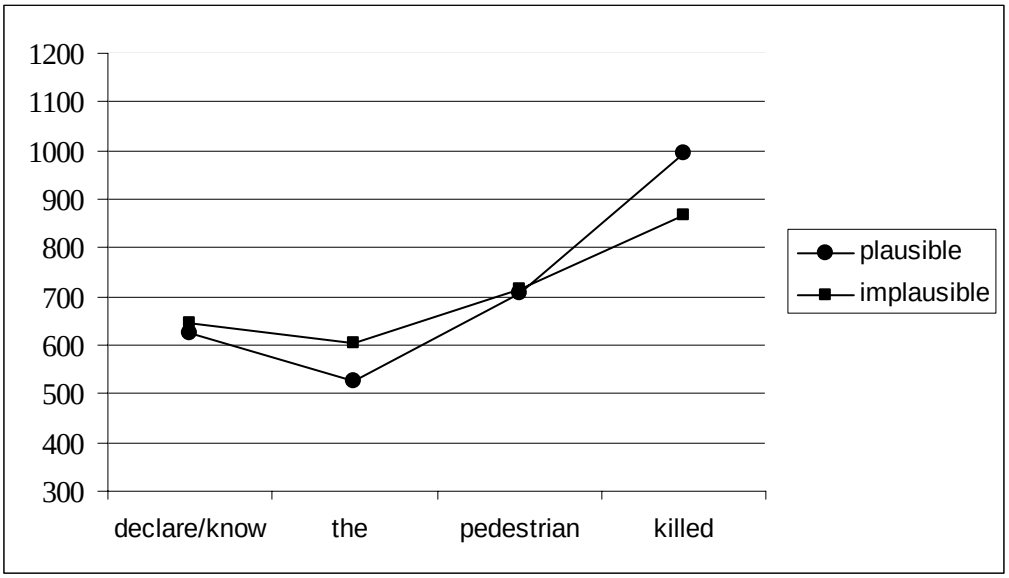

700
600
500
400
300
200
100
0
declare/know
killed
the
pedestrian
plausible
implausible


700
600
500
400
300
200
100
0
declare/know
the
pedestrian
killed
plausible
implausible


1200
1100
1000
900
800
700
600
500
400
300
declare/know
killed
the
pedestrian
plausible
implausible


1200
1100
1000
900
800
700
600
500
400
300
declare/know
the
pedestrian
killed
plausible
implausible


1200
1100
1000
900
800
700
600
500
400
300
declare/know
killed
the
pedestrian
plausible
implausible


1200
1100
1000
900
800
700
600
500
400
300
declare/know
the
pedestrian
killed
plausible
implausible

Appendix A
Experimental stimuli

Subject extraction (implausible)	Who did the client admit shot the banker?
Object extraction (implausible)	Who did the client admit the banker shot?
Subject extraction (plausible)	Who did the client mention shot the banker?
Object extraction (plausible)	Who did the client mention the banker shot?
Subject extraction (implausible)	Who does your mother agree amuses your friend?
Object extraction (implausible)	Who does your mother agree your friend amuses?
Subject extraction (plausible)	Who does your mother doubt amuses your friend?
Object extraction (plausible)	Who does your mother doubt your friend amuses?
Subject extraction (implausible)	Who does the police assert saw the killer?
Object extraction (implausible)	Who does the police assert the killer saw?
Subject extraction (plausible)	Who does the police fear saw the killer?
Object extraction (plausible)	Who does the police fear the killer saw?
Subject extraction (implausible)	Who did the director assume trained the assistant?
Object extraction (implausible)	Who did the director assume the assistant trained?
Subject extraction (plausible)	Who did the director hear trained the assistant?
Object extraction (plausible)	Who did the director hear the assistant trained?
Subject extraction (implausible)	Who did the coach say beat the champion?
Object extraction (implausible)	Who did the coach say the champion beat?
Subject extraction (plausible)	Who did the coach believe beat the champion?
Object extraction (plausible)	Who did the coach believe the champion beat?
Subject extraction (implausible)	Who did the detective claim kidnapped the scientist?
Object extraction (implausible)	Who did the detective claim the scientist kidnapped?
Subject extraction (plausible)	Who did the detective suspect kidnapped the scientist?
Object extraction (plausible)	Who did the detective suspect the scientist kidnapped?
Subject extraction (implausible)	Who did the principal conclude annoyed the student?
Object extraction (implausible)	Who did the principal conclude the student annoyed?
Subject extraction (plausible)	Who did the principal remember annoyed the student?
Object extraction (plausible)	Who did the principal remember the student annoyed?
Subject extraction (implausible)	Who did the attorney confirm misled the witness?
Object extraction (implausible)	Who did the attorney confirm the witness misled?
Subject extraction (plausible)	Who did the attorney mention misled the witness?
Object extraction (plausible)	Who did the attorney mention the witness misled?

Subject extraction (implausible)	Who did the Olympic judges decide surpassed the Russian gymnast?
Object extraction (implausible)	Who did the Olympic judges decide the Russian gymnast surpassed?
Subject extraction (plausible)	Who did the Olympic judges forget surpassed the Russian gymnast?
Object extraction (plausible)	Who did the Olympic judges forget the Russian gymnast surpassed?
Subject extraction (implausible)	Who did the police declare killed the pedestrian?
Object extraction (implausible)	Who did the police declare the pedestrian killed?
Subject extraction (plausible)	Who did the police know killed the pedestrian?
Object extraction (plausible)	Who did the police know the pedestrian killed?
Subject extraction (implausible)	Who did the CIA director deny saw the spy?
Object extraction (implausible)	Who did the CIA director deny the spy saw?
Subject extraction (plausible)	Who did the CIA director guess saw the spy?
Object extraction (plausible)	Who did the CIA director guess the spy saw?
Subject extraction (implausible)	Who did the attorneys dispute terrorized the man?
Object extraction (implausible)	Who did the attorneys dispute the man terrorized?
Subject extraction (plausible)	Who did the attorneys accept terrorized the man?
Object extraction (plausible)	Who did the attorneys accept the man terrorized?
Subject extraction (implausible)	Who did the nurse assume lied to the patient?
Object extraction (implausible)	Who did the nurse assume the patient lied to?
Subject extraction (plausible)	Who did the nurse overhear lied to the patient?
Object extraction (plausible)	Who did the nurse overhear the patient lied to?
Subject extraction (implausible)	Who did the mayor figure blackmailed the governor?
Object extraction (implausible)	Who did the mayor figure the governor blackmailed?
Subject extraction (plausible)	Who did the mayor conceal blackmailed the governor?
Object extraction (plausible)	Who did the mayor conceal the governor blackmailed?
Subject extraction (implausible)	Who did the secret agent hope talked to the prisoner?
Object extraction (implausible)	Who did the secret agent hope the prisoner talked to?
Subject extraction (plausible)	Who did the secret agent check talked to the prisoner?
Object extraction (plausible)	Who did the secret agent check the prisoner talked to?
Subject extraction (implausible)	Who did the jury infer murdered the athlete?
Object extraction (implausible)	Who did the jury infer the athlete murdered?
Subject extraction (plausible)	Who did the jury disclose murdered the athlete?
Object extraction (plausible)	Who did the jury disclose the athlete murdered?

Subject extraction (implausible)	Who did the editor insist pressured the writer?
Object extraction (implausible)	Who did the editor insist the writer pressured?
Subject extraction (plausible)	Who did the editor noticed pressured the writer?
Object extraction (plausible)	Who did the editor noticed the writer pressured?
Subject extraction (implausible)	Who did your mother maintain embarrassed the teacher?
Object extraction (implausible)	Who did your mother maintain the teacher embarrassed?
Subject extraction (plausible)	Who did your mother remember embarrassed the teacher?
Object extraction (plausible)	Who did your mother remember the teacher embarrassed?
Subject extraction (implausible)	Who did the nurse affirm visited the woman?
Object extraction (implausible)	Who did the nurse affirm the woman visited?
Subject extraction (plausible)	Who did the nurse observe visited the woman?
Object extraction (plausible)	Who did the nurse observe the woman visited?
Subject extraction (implausible)	Who did your sister wish married her friend?
Object extraction (implausible)	Who did your sister wish her friend married?
Subject extraction (plausible)	Who did your sister announce married her friend?
Object extraction (plausible)	Who did your sister announce her friend married?
Subject extraction (implausible)	Who did the teacher argue kissed the boy?
Object extraction (implausible)	Who did the teacher argue the boy kissed?
Subject extraction (plausible)	Who did the teacher recognized kissed the boy?
Object extraction (plausible)	Who did the teacher recognize the boy kissed?
Subject extraction (implausible)	Who did the editor conclude annoyed the reader?
Object extraction (implausible)	Who did the editor conclude the reader annoyed?
Subject extraction (plausible)	Who did the editor acknowledge annoyed the reader?
Object extraction (plausible)	Who did the editor acknowledge the reader annoyed?
Subject extraction (implausible)	Who did the show host assume insulted the contestant?
Object extraction (implausible)	Who did the show host assume the contestant insulted?
Subject extraction (plausible)	Who did the show host worried insulted the contestant?
Object extraction (plausible)	Who did the show host worried the contestant insulted?
Subject extraction (implausible)	Who did the investigator declare killed the hunter?
Object extraction (implausible)	Who did the investigator declare the hunter killed?
Subject extraction (plausible)	Who did the investigator discover killed the hunter?
Object extraction (plausible)	Who did the investigator discover the hunter killed?
Subject extraction (implausible)	Who did the hostess think poisoned the cook?
Object extraction (implausible)	Who did the hostess think the cook poisoned?
Subject extraction (plausible)	Who did the hostess fear poisoned the cook?
Object extraction (plausible)	Who did the hostess fear the cook poisoned?
Subject extraction (implausible)	Who does the teacher sense confused the student?
Object extraction (implausible)	Who does the teacher sense the student confused?
Subject extraction (plausible)	Who does the teacher know confused the student?
Object extraction (plausible)	Who does the teacher know the student confused?

Subject extraction (implausible)	Who did the man say hit the soccer player?
Object extraction (implausible)	Who did the man say the soccer player hit?
Subject extraction (plausible)	Who did the man believe hit the soccer player?
Object extraction (plausible)	Who did the man believe the soccer player hit?
Subject extraction (implausible)	Who did the press assure crushed the actress?
Object extraction (implausible)	Who did the press assure the actress crushed?
Subject extraction (plausible)	Who did the press announced crushed the actress?
Object extraction (plausible)	Who did the press announced the actress crushed?
Subject extraction (implausible)	Who did the reported write kidnapped the anchorman?
Object extraction (implausible)	Who did the reported write the anchorman kidnapped?
Subject extraction (plausible)	Who did the reported mention kidnapped the anchorman?
Object extraction (plausible)	Who did the reported mention the anchorman kidnapped?
Subject extraction (implausible)	Who did the producer figure entertained the comedian?
Object extraction (implausible)	Who did the producer figure the comedian entertained?
Subject extraction (plausible)	Who did the producer hear entertained the comedian?
Object extraction (plausible)	Who did the producer hear the comedian entertained?
Subject extraction (implausible)	Who did the store attendant hope followed the girl?
Object extraction (implausible)	Who did the store attendant hope the girl followed?
Subject extraction (plausible)	Who did the store attendant observe followed the girl?
Object extraction (plausible)	Who did the man observe the girl followed?
Subject extraction (implausible)	Who did the painted maintain chased the children?
Object extraction (implausible)	Who did the painter maintain the children chased?
Subject extraction (plausible)	Who did the painter imagine chased the children?
Object extraction (plausible)	Who did the painter imagine the children chased?