Article:
ROBERTS, LEAH and FELSER, CLAUDIA (2011) Plausibility and recovery from garden paths in second language sentence processing. Applied Psycholinguistics. pp. 299-331. ISSN: 1469-1817

https://doi.org/10.1017/S0142716410000421

Plausibility and Recovery from Garden Paths in Second Language Sentence Processing

Leah Roberts and Claudia Felser

Max Planck Institute for Psycholinguistics and University of Essex

Accepted for publication, *Applied Psycholinguistics*

Corresponding author:

Leah Roberts (leah.roberts@mpi.nl)

Max Planck Institute for Psycholinguistics

PO Box 310

6500 AH Nijmegen

The Netherlands

Abstract

In this study, the influence of plausibility information on the real-time processing of locally ambiguous ('garden-path') sentences in a nonnative language is investigated. Using self-paced reading, we examined how advanced Greek-speaking learners of English and native speaker controls read sentences containing temporary subject-object ambiguities, with the ambiguous noun phrase being either semantically plausible or implausible as the direct object of the immediately preceding verb. Besides providing evidence for incremental interpretation in L2 processing, our results indicate that the learners were more strongly influenced by plausibility information than the native speaker controls in their online processing of the experimental items. For the L2 learners an initially plausible direct object interpretation lead to increased reanalysis difficulty in 'weak' garden-path sentences where the required reanalysis did not interrupt the current thematic processing domain. No such evidence of online recovery was observed, on the other hand, for 'strong' garden-path sentences that required more substantial revisions of the representation built thus far, suggesting that comprehension breakdown was more likely here.

Introduction

In this study, we investigate second language (L2) learners' real-time processing of temporarily ambiguous sentences. While proficient L2 learners may be perfectly capable of comprehending even structurally complex sentences in their L2, we still know relatively little about the moment-by-moment mental processes involved in nonnative language comprehension. Results from several L2 processing studies indicate that even highly advanced learners may under-use morphosyntactic information when processing the L2 input in real time, but without necessarily performing any worse than native speakers in offline judgment or comprehension tasks (see Clahsen & Felser, 2006, for review). By way of accounting for these findings, Clahsen & Felser hypothesized that nonnative readers or listeners may be able to compensate for their grammatical processing difficulties by relying more on non-structural cues to interpretation such as semantic or pragmatic fit, world knowledge, or probabilistic information during comprehension. This account of L2 processing predicts that L2 learners will be highly sensitive to semantic information such as plausibility in their on-line processing, and this is what we examine in the current study.

We focus on L2 learners' processing of so-called 'garden-path' (GP) sentences, where much monolingual research has shown that the attempt at instant interpretation is likely to lead to an erroneous first analysis. For example, during the processing of sentences such as *While the band played the song pleased everyone*, the noun phrase (NP) *the song* is likely to be initially interpreted as the direct object of the preceding verb *play*, as indicated in (1a) below. However, the subsequent material does not fit with a direct object analysis as the main verb *pleased* requires an overt subject, forcing comprehenders to revise their initial analysis. Models of sentence comprehension differ in how they characterize the processor's preference for one analysis over another when there is more than one available, and in whether this preference results in the adoption of a single analysis to be pursued (compare e.g. Frazier &

Rayner, 1982) or a list of ranked alternatives (e.g. Gibson, 1991) or differing strengths of activation (e.g. McRae, Spivey-Knowlton & Tanenhaus,1998). Models agree, however, that when subsequent material is not compatible with the current preferred analysis, this analysis must be changed. This change can be envisaged as a re-ranking (e.g. Gibson, 1991; Spivey & Tanenhaus, 1998), competition between analyses (e.g., McRae et al., 1998; Tabor & Tanenhaus, 1999), pruning (e.g. Trueswell & Kim, 1998) or of entropy reduction of alternative analyses (Hale, 2003; see e.g. Vasishth & Lewis, 2006, for an overview). According to repair-based approaches (e.g. Frazier & Clifton, 1998; Gorrell, 1995; Sturt, Pickering, & Crocker, 1999) the changes required include detaching the NP *the song* from the verb phrase headed by *played* and reanalyzing it as a main clause subject, as shown in (1b) below, and the corresponding reassignment of case and thematic role features.

(1) a. [$_S$ While the band [$_{VP}$ played [$_{NP}$ the song]]] …

b. [$_S$ While the band [$_{VP}$ played __]] [$_S$ [$_{NP}$ the song] [$_{VP}$ pleased everyone]]

|_____↑

An increase in processing difficulty (compared to an unambiguous control condition) at or following the disambiguating element is commonly referred to as a 'garden-path effect'. Many studies have investigated what factors determine the relative cost of reanalysis (compare e.g. Fodor & Ferreira, 1998). For instance, semantic factors have been found to affect reanalysis (or re-ranking of a favored analysis) such that it appears to be more costly to the parser the more strongly committed the reader was to their initial analysis. That is, if a sentence fragment has been fully semantically interpreted and integrated into the current discourse model, an erroneous analysis will be more difficult to recover from (Frazier & Clifton, 1998). In some cases, native speakers may find themselves unable to revise a

strongly plausible initial misanalysis even in the face of contradictory grammatical evidence (Christianson, Hollingworth, Halliwell, & Ferreira, 2001).

Critical to our research question are the results of studies that have shown that the recovery process is affected by the plausibility of the ambiguous NP as a direct object of the preceding verb (Clifton, 1993; Stowe, 1989). For example, in a series of eye-movement monitoring experiments, Pickering and Traxler (1998) investigated how the plausibility of an initially favored analysis affects native English speakers' processing of sentences such as (2) below.

(2) As the woman edited (sailed) the magazine about fishing amused all the reporters.

The NP *the magazine about fishing* is temporarily ambiguous in that it can be analyzed either as the direct object of the preceding verb, or as the subject of the following main clause. Total reading times showed that the readers initially found the 'implausible' conditions more difficult to process because the direct object analysis here leads to a semantically implausible sentence fragment (i.e., #*the woman sailed the magazine*). Readers' subsequent recovery from this misanalysis was shown to differ as a function of the plausibility of the NP as a direct object of the preceding verb. In the region following syntactic disambiguation, more difficulty was visible in the processing of sentences containing plausible direct objects. This suggests that readers were more strongly committed to an initially plausible analysis than to an implausible one, and thus less willing to abandon it in the face of new evidence. We have adapted Pickering and Traxler's (1998) materials in the current experiment to examine the effect of plausibility on L2 learners' on-line processing commitments.

While Pickering and Traxler (1998) obtained similar results for temporarily ambiguous complement clause constructions such as (3) below, results from other studies suggest that this kind of sentence is easier to process than sentences such as (2) and does not necessarily lead to conscious GP effects (compare e.g. Holmes, Kennedy & Murray, 1987; Kennedy, Murray, Jennings, & Reid, 1989; Sturt, Pickering, & Crocker, 1999).

(3) The reporter saw her friend (her method) was not succeeding.

Holmes, Stowe and Cupples (1989), for example, failed to find any effects of the plausibility of the initial direct object (DO) analysis on readers' processing of the disambiguating auxiliary *was* in sentence like (3), suggesting that plausibility information did not affect reanalysis difficulty in GP sentences of this type. Questions regarding the details of the underlying mechanisms aside, most approaches to reanalysis agree that reinterpreting the ambiguous NP as a complement clause subject in 'weak' GP sentences like (3) requires less processing effort than elevating it to a main clause subject in structures of type (2). Those assuming a repair-type approach to reanalysis assume that this difference in cost arises because the latter involves reanalysis outside the current thematic domain (Gorrell, 1995; Pritchett, 1992; Sturt & Crocker, 1996) or command path (Weinberg, 1999).[1]

In sum, the present study builds on and extends previous research on the processing of temporary subject-object ambiguities by examining the extent to which plausibility information influences nonnative readers' ability to recover from an initial misanalysis in 'weak' versus 'strong' GP sentences. Before we present the results of the study, below we outline previous research on L2 learners' processing of garden-path constructions.

L2 learners' processing of garden-path sentences

While the number of studies investigating ambiguity resolution in nonnative language comprehension is still rather small (see Papadopoulou, 2005, and Frenck-Mestre, 2005, for reviews), there is evidence suggesting that L2 learners, like native speakers, try to interpret the input incrementally. Using a self-paced word-by-word grammaticality judgment task, Juffs and Harrington (1996) examined the processing of GP sentences by Chinese near-native speakers of English and native speaker controls. Their materials included sentences containing preposed adjunct clauses such as (4a,b) below.

(4) a. After Bill drank the water proved to be poisoned.

b. After Sam arrived the guests began to eat and drink.

While sentences such as (4a) that contain an optionally transitive verb usually elicit measurable GP effects in native speakers, such effects are expected to be reduced or absent in sentences such as (4b) that contain an intransitive verb, although some processing difficulty has been observed even with intransitive verbs (Van Gompel & Pickering, 2001). In addition, the authors tested GP sentences containing complement-clause ambiguities such as (5) below, which have been found to cause no or attenuated GP effects in monolingual processing.

(5) Sam warned the student cheated on the exam.

Juffs and Harrington found that both the learners and native speakers took longer to judge GP sentences than non-GP sentences and also had similar difficulties accepting GP sentences such as (4a) and (5) as grammatical. The analysis of the participants' word-by-word reading times showed that the L2 learners patterned with the native speakers in that both groups showed elevated reading times on the disambiguating verbs in the GP sentences.

This was taken to indicate that all participants initially analyzed the ambiguous NP as the direct object of the immediately preceding verb and were later forced to revise this analysis on encountering the disambiguating verb (i.e. *proved* in (4a), and *cheated* in (5)).

Although the authors did not set out to investigate factors that might affect recovery from misanalysis, the underlying assumption of this study is that the GP sentences should elicit false negative judgments more often than the non-GP control sentences because the built-in garden-path may lead to ultimate processing break-down. The error rates for both the L2 learners and the native speakers were indeed very high, with more than half of the GP sentences being incorrectly judged as 'impossible'. This could be taken to indicate that in many cases, neither the native speakers nor the L2 learners were able to recover from an initial misanalysis.

However, it is difficult to assess these results fully for two reasons. Firstly, readers were asked to provide acceptability judgments instead of being required to read for meaning, a task that may increase the likelihood of the experimental sentences being rejected for other reasons (an item may be deemed stylistically odd, for instance). Secondly, some extremely difficult GP items were included in the materials (e.g. *Ann convinced her friends were unreliable*), which may well account for the fact that in this condition, the native speakers achieved an accuracy score of only 22% which was even lower than the L2 groups' score of 35%. This finding is interpreted as evidence of the L2 learners' relatively greater ability to recover from garden-paths in these constructions, but the scores are so low that it is unlikely that many of these items were understood at all.

Juffs (2004) carried out a replication study examining how Chinese, Japanese and Spanish-speaking learners of English processed GP sentences of the above two types. While reanalysis of the postverbal NP is obligatory in English since finite clauses require an overt subject, Juffs hypothesized that sentences containing preposed adjunct clauses such as (4a)

might not cause particularly strong GP effects for speakers of null subject languages such as those above, in which an initial object analysis need not necessarily be corrected as the (phonetically empty) main clause subject can be discourse-identified. This prediction was not borne out, however. Despite this and other typological L1/L2 differences, the learners' reading patterns for GP sentences did not differ much from the native speakers'. All participants showed evidence of being garden-pathed in sentences of type (4a) - in the shape of elevated reading times at the disambiguating verb - and to a somewhat lesser degree when reading sentences of type (5). There was no evidence of any GP-induced processing breakdown in the grammaticality judgment data. Although the learners scored slightly worse than the native speakers overall, participants' error rates were considerably lower compared to what was seen in Juffs & Harrington's (1996) study, and they were not modulated by sentence type.

While the above findings suggest that nonnative readers may be garden-pathed to at least the same extent as native readers, relatively little is known about the role of semantic or pragmatic cues to interpretation in L2 ambiguity resolution. Williams, Möbius and Kim (2001) investigated the online use of plausibility information in L2 learners' processing of temporarily ambiguous sentences containing *wh*-dependencies. L2 learners of English from different L1 backgrounds (Chinese, Korean and German) read sentences such as (6) below that differed in the plausibility of the fronted *wh*-phrase as the direct object of the main verb (e.g. *push*).

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

Participants were asked to read the sentences one word at a time, and to indicate the point at which they thought the sentence had stopped making sense. Both the learners and the

native speakers made more 'stop' decisions at the verb in the condition containing implausible NPs than in the corresponding plausible condition, leading the authors to conclude that L2 learners, like native speakers, initially attempted to link a dislocated item to the first available potential subcategorizer (i.e., the verb *push*). Moreover, the postverbal disambiguation region in the implausible NP condition was easier to process compared to the same region in the plausible NP condition, although this effect was slightly delayed in the learners. A corresponding offline judgment task showed that the L2 learners judged the 'plausible' NP items as unacceptable more often than the 'implausible' items, despite the fact that by the end of the sentence, all experimental items were in fact fully grammatical. This could be taken to indicate that the learners had more difficulty than native speakers recovering from initial syntactic misanalysis, in line with Juffs and Harrington's (1996) findings. Note, however, that the learners' sensitivity to plausibility information during L2 reading may have been boosted by the fact that they were explicitly required to make use of this type of information here.

A follow-up study by Williams (2006) addressed the possibility of such task effects by adding a variant of the above experiment that involved a memory task instead of plausibility judgments. Interestingly, significant effects of plausibility were observed only where participants were explicitly required to provide online plausibility judgments but not in the memory task version of the experiment. Again, the learners patterned essentially alike irrespective of L1 background. Together, Williams et al.'s (2001) and Williams (2006) results show that when explicitly asked to pay attention to meaning, both native and nonnative readers are sensitive to plausibility information during processing, which in turn indicates that semantic or conceptual representations are built incrementally in nonnative language processing as well.

Taken together, the results form the above studies provide evidence both for the existence of garden-path effects in L2 processing and for learners' sensitivity to plausibility information during L2 reading. Among the questions that need to be investigated further, however, are (i) whether L2 learners also show evidence of being garden-pathed even when reading only for comprehension[2], and (ii) the extent to which the plausibility of the initial misanalysis affects learners' ability to recover the correct interpretation.

The current study

To compare the degree to which native and nonnative readers are influenced by plausibility information in online ambiguity resolution, we tested participants' processing of two different types of GP sentence. We examined the processing of 'weak' GPs in complement clause constructions such as (7a), and 'strong' GPs involving preposed adjunct clauses such as (7b).[3]

(7) a. The inspector warned the boss (crimes) would destroy very many lives.

b. While the band played the song (beer) pleased all the customers.

Specifically, we ask whether advanced Greek-speaking L2 learners of English find constructions with implausible initial analyses easier to recover from than those with plausible initial analyses, as has been found in earlier monolingual processing studies (Pickering & Traxler, 1998; Stowe, 1989; Clifton, 1993). If both native and nonnative readers evaluate the plausibility of a new incoming NP as a direct object of the preceding verb as soon as the NP is encountered, then we expect both participant groups to show evidence of initial misanalysis, reflected in elevated reading times for 'implausible' compared to 'plausible' NPs, in both types of GP sentence. If, however, learners are more strongly guided by plausibility information during L2 processing than native speakers, then effects of

plausibility should be more pronounced and/or prolonged for the learners, who should also have more difficulty than native speakers recovering from an initially plausible misanalysis. Finally, it is also possible that L2 learners are less sensitive to such 'higher-level' plausibility information than native speakers and as a result may rely more on structural processing strategies than native speakers.[4] In this case, for the L2 learners, initial misanalysis should occur following both 'plausible' and 'implausible' NPs, but there should be no differences in the recovery process between the two plausibility conditions. In the following, we set out the method and the results of this study.

The 'weak' and 'strong' garden-path items were presented in the same experiment, acting as fillers for each other, but for ease of exposition and because we cannot directly compare the two construction-types statistically, we will present the description of the materials and the results separately for each construction type.

Method

Participants

25 Greek speaking L2 learners of English (mean age: 24), and a control group of 24 native English speakers (mean age: 24) participated in the study. All participants had normal hearing and normal or corrected-to-normal vision. All were recruited from the student population of the University of Essex and were paid a small fee for their participation. No participant was informed of the ultimate purpose of the main experiment. The Greek participants had been resident in the UK at the time of testing for an average of 2.3 years and had first been exposed to English in a formal setting on average at age 8. None reported having acquired English along with Greek as a child, or having had any early naturalistic exposure to English. The mean English proficiency score of the Greek participants was 176 out of a possible 200 obtained from the Oxford Placement Test (Allan, 1992). This set the

group in the 'advanced' L2 learner category according to the Oxford Placement Test criteria. Table 1 summarises the Greek participants' bio-data and proficiency scores in English.

Insert Table 1 about here

Construction-type 1 - 'Weak' garden-paths

We firstly present the results of the processing of 'weak' GP sentences in complement clause constructions such as (7a) above.

Materials. The experimental materials consisted of 20 sentences such as (7a) above that contained optionally transitive verbs followed by a finite complement clause. Each experimental sentence was exactly 10 words long. The matrix verbs were selected from the materials used in earlier monolingual processing research (Pickering & Traxler, 1998)[5], and no verb appeared more than 3 times in the stimulus list. Only verbs whose selectional restrictions were comparable between Greek and English were used as matrix verbs so as to avoid any potential effects of L1 lexical transfer. The two experimental conditions were identical except for the critical NP, which was either plausible as a direct object of the preceding verb (e.g. *warned the boss*) or implausible (e.g. *#warned the crimes*).[6] The plausible and implausible NPs were matched for syllable length and frequency (according to the Celex Database, Centre for Lexical Information, 1993), and each noun appeared exactly once in a 'plausible DO' condition and once in an 'implausible DO' condition so as to control for possible but irrelevant effects of the critical noun itself. Thus each experimental list contained two instances of the same NP, once appearing as a plausible and once appearing as an implausible direct object.

To ensure that the experimental manipulation of plausibility worked as intended, 20 native English speakers (who did not participate further in the study) undertook an offline

plausibility rating task. They were presented with sentences such as *The inspector warned the boss* and *The inspector warned the crimes* and were asked to rate the plausibility of the constructions on a scale from 1 ('very plausible') to 7 ('very implausible'), and the results confirmed that the two conditions differed reliably in their plausibility scores (plausible mean = 1.2, SD = 0.3 versus implausible mean = 5.9, SD = 1.3; t (1, 19) = -16.79; $p < 0.001$). Each experimental item was followed by a yes/no comprehension question with the number of correct 'yes' and 'no' answers counterbalanced. Each question was formulated so as to ensure, as far as possible, that an accurate answer required the correct interpretation of an ambiguous NP as the complement clause subject, or proper recovery from any initial misanalysis (e.g. *The spokesman confirmed the story had surprised the president yesterday. Had the story surprised the president?)*. The experimental items were pseudo-randomized, mixed with the 20 'strong' garden-path items described below, and 42 additional fillers of different structural types. 10 of these fillers were 'pseudo-experimental' sentences, where an initial direct object analysis would in fact prove to be correct, such as *The woman suggested the idea to her husband at dinner*. These sentences were included in an attempt to ensure that the participants would not become sensitive to the experimental manipulation. Two thirds of the filler items were also followed by yes/no comprehension questions. Further to this, 8 practice items were included at the beginning of the experiment.

Procedure. Reading-time and comprehension accuracy data were collected using the non-cumulative moving-window procedure (Just, Carpenter, & Woolley, 1982). The participants were asked to read as quickly as possible on a 17" computer screen the set of experimental sentences and to answer any comprehension questions that followed an item. Both the experimental sentences and the accompanying questions were presented one word at a time, with participants bringing up each subsequent word by pushing a button on the push-button box. The final word of each sentence was indicated by a full stop. Participants

responded to the yes/no questions by pushing the corresponding buttons on the push-button box. After each trial, the experiment was temporarily suspended, which allowed the participants to pause for as long as they wanted to before they pressed a designated key on the computer keyboard to resume the session. The experiment was controlled and the results recorded by the software package NESU (Baumann, Nagengast, & Klaas, 1983) and took no longer than 40 minutes to complete, depending on how long each participant chose to rest between trials.

Results

Accuracy. Comprehension accuracy was high for both participant groups, with the Greek L2 learners answering correctly 93% and the native speakers 89% of the questions following the experimental items. Table 2 shows the groups' mean accuracy scores for each condition.

Insert Table 2 about here

An ANOVA was run on the arcsine transformed accuracy data with the within-subjects factor Plausibility (plausible DO, implausible DO) and the between-subjects factor Group (L1, L2). There was a significant interaction between Plausibility and Group ($F1\ (1, 47) = 7.06$; $p < 0.02$; $\eta^2 = .13$; $F2\ (1, 38) = 4.96$; $p < 0.03$; $\eta^2 = .12$), reflecting the fact that the learners were more accurate in responding to questions following the plausible DO condition compared to the implausible DO condition, whereas the opposite pattern was found in the native speakers[7], but in neither group was the comparison between conditions statistically significant (L1: $t1\ (23) = 1.93$; $p > .07$; $t2\ (19) = 1.53$; $p > .1$; L2: $t1\ (24) = 1.901$; $p > .07$; $t2\ (19) = 1.679$; $p > .1$).

Reading times. In order to assess fully the processing of the experimental items and the impact of the experimental manipulations, we examined the reading times (RTs) on each of the segments (which comprised single words) from the head noun of the ambiguous NP until the end of the sentence. Only those items to which the comprehension questions were correctly answered were analyzed. We also analysed the length of time it took participants to read through the comprehension questions that followed the experimental items. Prior to the analysis, individual RTs above a time-out of 3000 ms for participants were removed (English 0.03 %, Greek 1.31 %). Figure 1 shows the two participant groups' reading profiles for correct trials for each experimental condition.

Insert Figure 1 about here

In order to explore whether the learners' reading profiles differed reliably from the native speakers', we ran separate mixed ANOVAs on the mean RTs for each of the 6 segments including and following the postverbal noun with the within-subjects factor Plausibility (plausible DO, implausible DO) and the between-subjects factor Group (L1, L2). As the two experimental conditions were identical up to the point at which the ambiguous NP was presented, only the results for segments 5-10 will be reported. We further analysed the summed reading times of the comprehension questions which immediately followed the experimental items. Thus the analyses were run on 6 segments of the experimental sentences, and on the summed reading times of the comprehension questions. The results of these analyses are summarized in Table 3.

Insert Table 3 about here

The fact that the nonnative participants were slower, on average, to read the experimental sentences than the native speakers was reflected by the significant main effects of Group that were found on all segments.

Segment 5 contained the head noun of the ambiguous NP (*book*/*girl*). As well as a main effect of Plausibility, there was an interaction between Plausibility and Group that was significant by items and near-significant in the analysis by participants, indicating that the participant groups were affected differently by the plausibility manipulation. While the L2 learners' processing of the ambiguous NP was affected by its plausibility as the direct object of the preceding verb, with mean RTs to the implausible DO condition significantly higher than those to the plausible DO condition ($t1\ (24) = 2.07$; $p < 0.05$; $t2\ (19) = 2.31$; $p < 0.04$), there was no difference between the two conditions for the native speakers ($t1\ (23) = .399$; $p > 0.6$; $t2\ (19) = .252$; $p > 0.8$).

On the following segment (6), which contained the disambiguating auxiliary (e.g. *had*), there was a main effect of plausibility indicating that both groups now found the plausible DO condition easier to process than the implausible DO condition. Although this effect was numerically greater in the RTs of the L2 group (an 83 ms disadvantage difference for implausible items compared to 29ms for the native speakers), the lack of interaction with Group shows that there was no statistically reliable difference between the two groups. This plausibility effect was not as robust as on the previous segment, with t-test comparisons showing that the difference was statistically significant for the native speakers by subjects only ($t1\ (23) = 2.28$; $p < 0.04$; $t2\ (23) = 1.18$; $p = .255$) and marginally significant for the L2 learners ($t1\ (24) = 1.96$; $p = 0.061$; $t2\ (23) = 2.05$; $p = 0.054$).

On segment 7, there was a significant main effect of Plausibility by items. Now the plausible items elicited higher RTs than the implausible items for both groups, a reversal in

the pattern of RTs seen earlier, again, this comparison was numerically stronger in the learners (65 ms) than in the native speakers (13 ms), although again, there was no interaction with Group and no t-test comparison revealed statistically significant differences (p's > .1).

This reverse plausibility effect was more strongly in evidence in segment 8, were there was a significant Plausibility by Group interaction by participants. There was no difference between the two conditions in the native speaker data ($t1\ (22) = .249$; $p > .8$; $t2\ (19) = .719$; $p > .4$), whereas for L2 learners the plausible condition elicited significantly higher RTs than the implausible condition ($t1\ (24) = 2.38$; $p < 0.03$; $t2\ (19) = 2.24$; $p < 0.05$). There were no statistically reliable effects of plausibility on the penultimate segment 9 or segment 10.

The reverse plausibility effect was observed in the reading of the comprehension questions for both groups, with analyses on the summed RTs revealing a significant main effect of Plausibility by participants, but no interaction with Group. Both groups read the questions which followed the plausible DO items more slowly than those following the implausible items (L1s: 3642 ms vs. 3524 ms; L2s: 2982 ms vs. 2826 ms) suggesting that these comprehension questions more difficult than those following the implausible DO.

Summary. Whereas the L2 group showed an immediate and prolonged effect of plausibility during their reading of the ambiguous NP, this effect was delayed (and numerically weaker) in the native speakers. The reverse plausibility effect was observed on segments 7 for the learners and it continued on into segment 8, indicating that they subsequently found it more difficult to process the condition in which their earlier misanalysis had led to a plausible interpretation, and this effect was also in evidence in both groups' processing of the comprehension questions.

Construction-type 2 - 'Strong' garden-paths

Below we report the results of the participants' online processing of 'strong' garden-path constructions involving preposed adjunct clauses such as (7b).

Materials. The materials consisted of 20 sentence pairs containing preposed adjunct clauses such as *While the band played the song (beer) pleased all the customers*, constructed from 14 optionally transitive verbs (comparable with Greek on selectional restrictions) that were used no more than 3 times each. Each critical noun (matched for syllable number and frequency) appeared once in a plausible DO condition and once in an implausible DO condition. As with the complement clause items, 20 native English speakers who did not take part in the main experiment rated the plausibility of each experimental pairing of verb and potential direct object. The analysis of these ratings showed that the 'plausible' items were indeed considered more plausible than the 'implausible' items (plausible mean = 1.5, SD = 0.7 versus implausible mean = 6.2, SD = 0.7; $t(19) = 20.56$; $p < 0.001$) confirming that our experimental manipulation worked as intended.

The experimental items were all followed by a counterbalanced number of yes/no comprehension questions and mixed with the 'weak' GP sentences and 42 fillers, a third of which was also followed by a question (see Appendix). Included in the items were 10 'pseudo-experimental' items, where an initial direct object analysis for an ambiguous NP is in fact correct (e.g. *While the president ate his lunch a bodyguard stood by*).

Accuracy. Again, participants scored highly on the comprehension measure (Greek: 91%, English: 94%). Table 4 shows the two groups' mean accuracy scores on the comprehension questions following the experimental items.

Insert Table 4 about here

An ANOVA, again run on the arcsine transformed accuracy data, with the within-subjects factor Plausibility and the between-subjects factor Group which was run on the mean accuracy scores revealed a significant main effect of Plausibility (F1 (1, 47) = 7.69; $p < 0.01$; $\eta^2 = .14$; F2 (1, 38) = 5.71; $p < 0.05$; $\eta^2 = .13$) and a Plausibility by Group interaction (F1 (1, 47) = 15.17; $p < 0.001$; $\eta^2 = .24$; F2 (1, 38) = 7.83; $p < 0.008$; $\eta^2 = .17$). This was due to the learners' answering the questions following 'plausible DO' items significantly less accurately than those following 'implausible DO' sentences (t1 (23) = -4.11; $p < 0.001$; t2 (19) = -3.21; $p < 0.01$), whereas the native speaker's accuracy scores did not differ significantly between the two conditions (t1 (23) = .982; $p > 0.3$; t2 (19) = .346; $p > 0.7$).

Reading Times. As before, only the reading times from correct trials were analyzed. The same data cleaning procedure was used as with the 'weak' GP constructions, with the removal of time-outs affecting 1 % of the Greek and 0.08 % of the native speakers' data. Figure 2 shows the participants' reading profiles for both experimental conditions (correct trials only).

Insert Figure 2 about here

To see whether there were any differences in the two groups' processing of the experimental items from the ambiguous NP onwards, mixed ANOVAs were run with the within-subjects factor Plausibility and Group as a between-subjects factor per segment of the experimental items from the head of the ambiguous NP (Segment 6) and also on the summed RTs of the comprehension questions which followed the experimental items. The results are summarized in Table 5.

Insert Table 5 about here

As with the 'weak' GPs, overall the learners read the experimental sentences more slowly than the native speakers, and this was supported by the significant main effects of Group across all segments. Further to this, there was a significant interaction between Plausibility and Group on Segment 6, which was marginal by items on Segment 7 indicating differences in the two participant groups' processing of the experimental items. On Segment 6, where the head noun of the ambiguous NP was presented, 'plausible' items elicited shorter RTs than 'implausible' ones for the Greek speakers ($t1\ (24) = -2.16$; $p < 0.05$; $t2\ (19) = -3.46$; $p < 0.01$) but not for the native speaker controls ($t1\ (23) = .181$; $p > 0.8$; $t2\ (19) = .067$; $p > 0.9$). Similarly to their processing of the complement clause GPs, the processing difficulty that the L2 learners displayed for the implausible versus the plausible DO condition was carried over into the following disambiguating segment (i.e., *pleased*) ($t1\ (24) = -2.14$; $p < 0.05$; $t2\ (19) = -1.73$; $p > 0.09$), whereas there was no difference in the native speakers' RTs ($t1\ (23) = .388$; $p > 0.7$; $t2\ (19) = .486$; $p > 0.6$). There was a marginally significant main effect of Plausibility on Segment 9, because the plausible condition elicited higher RTs than the implausible condition, but this was not statistically robust. No significant effects or interactions (other than main effects of Group) were found on Segments 8 or 10, nor on the comprehension question reading times.

Summary. The L2 learners showed a prolonged plausibility effect, as they did with the complement clause items, beginning at the head noun of the ambiguous NP and continuing into the immediately following segment where the disambiguating verb was presented, with the implausible DO condition causing significantly higher RTs than the plausible DO condition. Unlike in their processing of the complement clause items, however, there was no statistically robust evidence of a reversal of this effect on any of the later segments, nor on their reading times for the comprehension questions. The analysis of the comprehension

accuracy scores, however, showed that the L2 learners made significantly more errors in the plausible DO condition compared to the implausible DO condition. There was no statistical difference in the native speakers' processing of the experimental items, suggesting that plausibility information had no measurable effects on their processing of this type of GP sentence.

Post hoc analyses: Reading Speed. Even though we had deliberately kept our stimulus materials reasonably simple in an attempt to make them easier for L2 learners to comprehend, it is puzzling that the native speakers in this experiment failed to show (robust) plausibility effects for either the 'weak' or the 'strong' garden-path sentences, given earlier findings from monolingual research (e.g. Pickering and Traxler, 1998). One possibility[8] is that the native speakers were reading the experimental items so quickly that such plausibility effects may have been masked. Therefore, we investigated the potential effect of overall reading speed on the processing of the experimental items in a post-hoc analysis. For each group separately, we entered the participant's mean reading speed (computed across all experimental items, fillers and questions) as a covariate factor into an ANOVA, together with the within-subjects factor Plausibility. We performed these analyses both for the native speakers and the L2 learners, and on the data from both the 'weak' and the 'strong' garden-path constructions.

For the native speakers' processing of the 'weak' GP data, shown in Table 7, the speed analysis found a significant interaction between Reading Speed and Plausibility on Segment 5, the ambiguous noun, and Segment 8, and a marginal interaction on Segment 7 ($F(1, 22) = 3.26$; $p = 0.085$; $\eta^2 = .13$). As can be seen from the mean reading times for each speed group in Table 6 (following a median split—obtained from the median RTs of all regions—at 359.09 ms), the slower L1 readers spent more time reading the 'implausible' DO than the 'plausible' DO items on Segments 5 and 6. This pattern was reversed on Segments 7 and 8, where the plausible items elicit higher RTs, although t-test comparisons were only significant

on Segment 6 by subjects (t1 (11) = -2.35; $p < 0.05$; t2 (19) = -1.47; $p = 0.16$). Therefore, the effect of misanalysis observed in the native speakers as a whole group on Segment 6 in the main analysis reported above was clearly carried by this group of slower native speaker readers.

Insert Table 6 about here

Insert Table 7 about here

The reading speed analysis run on the native speakers' comprehension accuracy scores for 'weak' GPs also revealed a significant interaction between Plausibility and Reading Speed. In contrast to the reading time analysis, this interaction was now driven by the faster reading L1 group, whose accuracy was significantly lower on the comprehension questions following the plausible items (83%) compared to the implausible ones (94%) (t1 (11) = -2.35; $p < 0.05$; t2 (19) = -2.11; $p < 0.05$), whereas there was no significant difference in the accuracy scores of the slower reading group (88% vs. 90%, p's > .5). This suggests that the faster readers had difficulty recovering the correct interpretation for the plausible DO items, whereas the slower group had recovered from their initial misanalyses as observed in the RTs.

Reading speed also affected the native speakers' processing of the 'strong' GP items (see Tables 8 and 9) with a significant interaction between Plausibility and Reading Speed on Segment 7, the disambiguating verb, (Table 8 shows the mean RTs for each group and Table 9, the results of these analyses). On Segment 7, the two groups' RTs patterned in the opposite direction, with the slower group showing elevated RTs for the plausible versus the implausible DO conditions, although no comparison was statistically significant (p's >.2). This effect carried over into the subsequent segments, with the 'plausible' items causing

significantly higher RTs than the 'implausible' ones for the slow L1 readers on Segment 9 (t1 (11) = 2.46; $p < 0.05$; t2 (19) = 2.12; $p < 0.05$) whereas there was no difference in RTs for the faster L1 readers on this segment (p's > .3). There were no effects of Reading Speed on either the mean RTs on the comprehension questions, nor on the accuracy scores. In sum, similarly to their processing of the 'weak' GP items, only the slower native speaker readers showed significant effects during the on-line processing of the experimental items.

Insert Table 8 about here

Insert Table 9 about here

We also ran a speed analysis on the L2 learners' data.[9] The results of these analysis are presented in Table 11. Their processing of the 'weak' GP items was affected by their overall speed on the segment containing the ambiguous NP, Segment 5 ($F (1, 23) = 9.51$; $p < 0.01$; $\eta^2 = .29$) and Segment 8 in the disambiguating region ($F (1, 23) = 8.80$; $p < 0.01$; $\eta^2 = .28$). We divided the L2 group according to the median reading speed (at 585.37 ms) and Table 10 shows the reading times for the faster (n = 13) and slower (n = 12) L2 readers.

Insert Table 10 about here

Insert Table 11 about here

Investigating the interactions on these two segments, it was found that although for both learner groups the plausible condition elicited higher RTs than the implausible condition on segment 5, it was much stronger, and statistically significant, only in the slower L2 readers (t1 (11) = -2.20; $p < .05$; t2 (19) = -2.35; $p < 0.05$ vs. t1 (12) = -.515; $p > 0.7$; t2 (19) = -.158;

p > 0.8). The interaction with Reading Speed on Segment 8 was caused by the slower L2 group who spent significantly more time reading the plausible than the implausible condition (690 vs. 573: t1 (11) = 3.41; p < 01; t2 (19) = 2.64; p < 05), as they did on the previous segment 7, whereas there was no difference for the faster L2 group on this spill-over segment. This prolonged effect suggests that the slower L2 group (who were also less proficient) found it more difficult to recover from their initial misanalysis.

For the 'strong' GP data (see Table 13), the L2 learners' processing of the ambiguous NP, Segment 6, was affected by their mean reading speed (F (1, 23) = 7.51; p < 0.05; η^2 = .25) such that the slower L2 group had significantly more processing difficulty with the implausible than the plausible condition at this point (t1 (11) = 2.66; p < .05; t2 (19) = 3.86; p < 0.01), whereas there was no difference in RTs for the fast L2 readers on this segment (p's > 0.9) (see Table 12). There was also a marginally significant interaction on Segment 9. Numerically, for both L2 groups, there appeared to be a reverse plausibility pattern in the RTs. This effect was greater and marginally significant by participants only for the faster L2 group (31 ms vs. 11 ms) (t1 (12) = 2.04; p = 0.064; t2 (19) = 1.34; p = 0.196). There were no other differences between the two reading speed L2 groups' RT. In sum, as in their processing of the 'weak' GPs, the slower L2 readers showed an immediate and prolonged effect of the postverbal NP's plausibility as a direct object. This effect was also visible for the faster (and more proficient) L2 group, but it was confined to one segment (Segment 7), and was not in evidence in the following segment (8). This group also showed a hint of recovery from their misanalysis in their RTs during the post-disambiguation region, which mirrored the reverse plausibility effect seen in the learners' accuracy scores when the whole group is taken together (Table 4). In sum, the plausibility of the initial misanalysis affected the L2 learners' recovery, but the effect appeared earlier in the faster L2 group, in evidence during their

processing of the experimental items, whereas both groups (when collapsed) showed this reverse plausibility effect in the pattern of their accuracy responses.

Insert Table 12 about here

Insert Table 13 about here

Discussion

The purpose of this study was to investigate how plausibility affects nonnative readers' processing of GP sentences and their ability to recover from initial misanalysis. In order to address this question, we presented participants with GP sentences containing either complement or preposed adjunct clauses and manipulated the temporarily ambiguous NP such that it was either plausible or implausible as the direct object of the initial verb. Our main findings can be summarized as follows:

- Implausible direct object NPs in 'weak' GP sentences led to longer RTs in both participant groups, with the effect of our plausibility manipulation being visible immediately and prolonged in the learners but numerically smaller and delayed in the native speakers.
- Later on during the processing of 'weak' GPs, the earlier plausibility effect reversed in the L2 group, indicating greater reanalysis difficulty for an initially plausible misanalysis than for an implausible one.
- For 'strong' GP sentences containing preposed adjunct clauses, implausible DOs elicited longer RTs than plausible ones only for the L2 group. This effect was again

immediate and prolonged, also being visible in the immediately following segment where the disambiguating verb was presented.

- While both participant groups showed high comprehension accuracy overall, the Greek group differed from the native speaker controls in that they showed significantly lower accuracy scores to 'strong' GPs containing plausible DOs compared to those containing implausible ones. The two group's accuracy scores to 'weak' GPs patterned in the opposite direction, with the native speakers' responding less accurately to those containing plausible DOs, but no comparisons for either group's accuracy data were statistically significant.
- The segment-by-segment processing of the experimental items was found to be affected by the overall reading speed for both the L2 learners and the native speakers. For the slower native speaker readers, there was evidence of initial misanalysis in weak GPs as well as reverse plausibility effects in strong GPs visible in their on-line processing of the experimental items, whereas there were no visible effects it the RTs of the faster L1 group.
- For the L2 learners, in general, the overall pattern found for the whole L2 group with both GP types did not differ dramatically from the pattern of RTs when the group was broken down into faster and slower readers. Although the effects were weaker overall in the faster L2s, the speed groups mainly differed in the start (later for the fast group) and the duration (longer for the slower group) of the plausibility effects observed for the whole group. The faster L2 learners (who were also the most proficient in English) also showed the beginnings of a reverse plausibility effect the disambiguating region of the 'strong' GPs with plausible DOs, which suggested that they attempted recovery from their initial misanalyses earlier than the slower L2 learners.

In the following, we will discuss our findings in more detail in the light of previous research and of the questions that were raised at the outset.

Incremental interpretation in L1 and L2 comprehension

In line with the findings from earlier L2 processing studies (e.g., Juffs & Harrington, 1996; Juffs, 2004; Williams et al., 2001; Williams, 2006), our Greek-speaking participants showed evidence of processing the input incrementally. The observed increase in reading times triggered by implausible compared to plausible DOs for both GP construction types indicates that the learners evaluated the ambiguous NP's plausibility as a direct object of the immediately preceding verb, and for the slower L2 readers, this effect occurred as soon as they encountered it. The fact that plausibility effects were in evidence for all L2 learners, and immediate and prolonged in the slower L2 group, but only in evidence for the slower native speaker readers among the native controls suggests that nonnative comprehenders may be influenced more strongly than native ones by pragmatic plausibility information during processing.[10] The findings go against the idea that L2 learners may be less sensitive to certain 'higher-level' plausibility information than native speakers, and as a result, rely more strongly on structural information. This is in line with the findings of other L2 studies, for instance, the plausibility effects observed in the learners in Williams et al.'s (2001) and Williams' (2006) studies, as well as with preliminary evidence suggesting that nonnative comprehenders are more sensitive to discourse-pragmatic information than native ones (Felser, Sato & Bertenshaw, 2009; Roberts, Gullberg & Indefrey, 2008).

There are several reasons why the native speakers may have been less troubled by our plausibility manipulation than the L2 learners. Although the experimental materials were based on those from a study undertaken by English monolinguals (Pickering & Traxler, 1998), we adapted them so that they would be appropriate for L2 learners, keeping them comparatively simple. Among the various factors that have been found to affect the relative

strength of garden-path effects in native speakers is the structural distance between the ambiguous NP and the disambiguating element (e.g. Ferreira & Henderson, 1991; Christianson et al., 2001). That is, increasing the distance between the head noun of the ambiguous NP and its true subcategorizer as, for example, in Pickering and Traxler's (1998) materials (e.g. *As the woman edited the magazine about fishing amused…*) might have increased reanalysis difficulty for their native speakers, because the parser is committed to the wrong analysis for longer. In the current materials, in contrast, the disambiguating verb or auxiliary appeared directly adjacent to the ambiguous NP in both types of GP sentence, and thus commitment to the wrong analysis would have been relatively short-lived.[11]

Note that in the absence of non-GP control conditions, our experiment was designed to measure the relative influence of plausibility information on participants' processing of GP sentences, rather than differences in the processing of structurally ambiguous versus unambiguous sentences (compare also Pickering & Traxler, 1989). If, as serial or 'syntax-first' models of sentence processing have claimed (e.g. Frazier, 1987), initial parsing decisions in the L1 are made on purely structural grounds irrespective of semantic or pragmatic fit, then manipulating the plausibility of the direct object analysis may not have been a sensitive enough diagnostic for initial misanalysis (or reanalysis difficulty) in the comparatively simple GP sentences that were used in the current study. In other words, the native speakers may have been able to correct their initial parsing error before having had a chance to commit strongly enough to the wrong interpretation for plausibility to have any measurable effects on reanalysis difficulty.[12]

However, the results of our post-hoc analysis found that individual differences in reading speed affected the online processing of these items. Only the slower native readers showed measurable evidence of evaluating the plausibility of the ambiguous postverbal NP as a direct object of the preceding verb when this NP was encountered, and evidence that their

subsequent recovery from the GP was affected by the plausibility of the original misanalysis. In contrast, the faster English readers either showed dramatically delayed effects in comparison to the slower readers (misinterpretation evident on the comprehension question accuracy only, in the 'weak' GP condition) or no evidence of misanalysis or online recovery, according to our 'plausibility' diagnostic, at all ('strong' GPs). At least in the former condition, this suggests that only the slower readers were likely to have been processing each word as they encountered it, whereas the faster group perhaps sped through the items, and *then* performed the relevant analyses while answering the comprehension questions. This would explain why no online processing costs were observed for the faster readers, but that they were significantly slower in reading the comprehension questions and significantly less accurate in answering them when they followed a plausible DO, at least for the complement clause items. Interestingly, this finding is suggestive of earlier monolingual processing work which has found that the semantic representation of earlier misanalyses may persist following readers processing of garden-path sentences, as evidenced by inaccurate responses to the comprehension questions that followed them (Christianson et al., 2001; Christianson, Williams, Zacks, & Ferreira, 2006).[13]

Recovery from garden paths in L2 sentence processing

As stated above, our results suggest that the plausibility of the ambiguous NP affected the participants' ability to recover from their initial misanalysis, but differently so for the two types of GP sentence under investigation. For the L2 learners, for both construction types there was a prolonged plausibility effect indicative of an initial attempt to interpret the ambiguous NP as a direct object of the preceding verb. A reversal of this effect was observed only for 'weak' GP sentences involving complement clauses but not for 'strong' GPs containing preposed adjunct clauses, however, when the L2 group was analysed as a whole.

With the complement clause items, the expected reversal, which is thought to reflect online reanalysis difficulty (compare e.g. Pickering & Traxler, 1998), was clearly visible in the post-disambiguation region for all L2 learners. This indicates that for items where the initial DO analysis resulted in a plausible interpretation, subsequent recovery from this misanalysis was more costly, suggesting a stronger commitment to the initial misanalysis if this was pragmatically plausible. The learners' high comprehension accuracy score for implausible DO items moreover suggests that recovery was usually successful. In the processing of the 'strong' GPs, in contrast, the absence of any such reversal during the processing of the remainder of the sentences for the slower, less proficient learners would seem to indicate that no revision of the initial analysis was attempted during the online processing of 'strong' GPs – despite the learners' obvious sensitivity to our plausibility manipulation during their processing of the ambiguous NP. The faster group's higher RTs on the plausible DO items in the region following disambiguation might suggest an attempt at recovery, although the observed numerical trend was not statistically reliable. However, the learners' relatively poorer comprehension accuracy scores for plausible DO items in this experiment moreover suggests that they had difficulty recovering the correct interpretation even after finishing reading the whole sentence, even the more highly proficient L2 group.

This difference in difficulty found in the L2 learner data in the recovery process between the two construction types is in line with earlier findings from the monolingual processing literature and may be accounted for by structural differences between the two. As stated earlier, models of sentence comprehension differ in how reanalysis is handled (e.g., Gibson, 1991; Hale, 2003; Levy, 2008; Spivey & Tanenhaus, 1998, Trueswell & Kim, 1998) but all assume that 'weak' and 'strong' GPs would elicit different processing costs. For instance, according to structurally-based parsing theories (e.g. Gorrell, 1995; Pritchett, 1992; Weinberg, 1999), the process of reanalyzing the ambiguous NP as a complement clause

subject in 'weak' GP sentences may involve relatively little computational effort because only minimal changes to the representation built thus far are required. In sentences like (7a), reanalysis involves adding a clause boundary (as shown in (8) below) and revising case and thematic role assignments to the NP *the book* (see also Gorrell, 1995; Sturt & Crocker, 1996). As these adjustments can all take place within the current thematic processing domain (Pritchett, 1992) or command path (Weinberg, 1999).

(8) a. The inspector $[_{VP}\ [_{V'}$ warned $[_{NP}$ the boss]]] …

b. The inspector $[_{VP}\ [_{V'}$ warned $[_{S}\ [_{NP}$ the boss] $[_{VP}$ would destroy …]]]]

The situation is different for GP sentences containing preposed adjunct clauses such as (7b). Upon encountering the disambiguating verb *pleased* in the sentence *While the band played the song pleased all the customers*, the processor needs to reanalyze the ambiguous NP as the subject of the superordinate clause (compare (1b), repeated below for convenience). This requires a change in the dominance relations built so far and the creation of a new thematic processing domain.[14]

(1) b. $[_{S}$ While the band $[_{VP}$ played __]] $[_{S}\ [_{NP}$ the song] $[_{VP}$ pleased everyone]]

|_____↑

The relative difficulty of undoing the initial misanalysis here may have been such that unlike with the 'weak' GPs, the learners as a whole group were unable to perform the required reanalysis, including the reassignment of thematic roles, during their processing of the remainder of this kind of GP sentence. There was a hint, however, that the more proficient of the L2 group started the recovery process during the reading of the experimental items.

That nonnative readers should have difficulty performing online reanalysis in 'strong' GP sentences is in fact unsurprising, given findings from earlier L2 processing studies showing that learners' syntactic processing abilities may be reduced, or less automatized, compared to native speakers' (Felser & Roberts, 2007; Felser, Roberts, Gross & Marinis, 2003; Hahne & Friederici, 2001; Marinis, Roberts, Felser & Clahsen, 2005; Papadopoulou & Clahsen, 2003). Our results indicate that whereas minor structural revisions and/or reassignment of thematic roles within the current thematic processing domain did not seem to pose much of a problem for our L2 group, more radical structural reorganization (involving a break in the current thematic domain or command path) may be difficult even for highly proficient learners to achieve online.

Another point of interest here is the clear difference in how the native speakers processed the two GP types in comparison to the L2 learners. The post-hoc speed analysis found that only the 'weak' GP constructions elicited online processing effects of both misanalysis and recovery in the native speakers (at least for the slower readers). Furthermore, reverse plausibility effects were observed in their reading of the comprehension questions, as well as in the faster readers' accuracy scores in these 'weak' GPs. In contrast, for the 'strong' GPs, there was only a robust reverse plausibility effect in evidence, and only for the slower readers. Unlike for the 'weak' GP items, as well as in contrast to the L2 learners in the 'strong' GPs, accuracy was high in both plausibility conditions. This suggests that all the native speakers were able eventually to recover the correct interpretation of the 'strong' GP items. This raises the question of why this type of GP sentence should apparently have been easier to process for the native speakers, given that GP sentences containing preposed adjunct clauses are usually assumed to trigger stronger GP effects than those containing complement clauses.

The most likely explanation for this, in our view, is that in our 'strong' GP items, the ambiguous NP was immediately followed by a disambiguating content verb, which provided both a syntactic and a semantic cue as to NP's correct interpretation. As a result, the native readers may have been able to revise any possible initial misanalysis too quickly for this to lead to any measurable effects of plausibility here – especially since the predicted effects go into opposite directions, so may have cancelled each other out.[15] This would also account for our observation that trends in the expected direction were observed only for slow – but not for relatively faster – native readers here. In the 'weak' GP items, in contrast, the segment following the ambiguous NP was a modal or auxiliary verb, which provided a syntactic disambiguation cue but no clear indication of the NPs correct thematic role. The lexical licenser (or thematic role assigner) of the embedded subject NP was not provided until the following segment or even later, which may have given participants more time for evaluating the plausibility of the NP as a direct object of the first verb, leading to more clearly visible recovery problems in the case of our 'weak' GP items.

If nonnative readers are even more strongly guided by semantic information than native speakers during L2 sentence processing, then the delayed presentation of the postverbal NP's actual lexical licenser in the 'weak' GP sentences might help account for our finding that despite showing initial sensitivity to our plausibility manipulation, the learners as a whole showed little evidence of online recovery.

Comprehension accuracy

Participants' ability to answer our end-of-trial comprehension questions can provide useful information about whether or not they ultimately managed to recover from being garden-pathed. Recall that in the current study participants were asked to read the sentences for meaning and subsequently answer questions that always targeted the ambiguity. On the

basis of results from earlier studies, it was predicted that the L2 learners would find interpreting GP sentences more difficult overall than the native speakers, indicated by lower accuracy scores. It is very striking, therefore, that this was not generally the case here. Overall, the accuracy rates did not differ between the two groups, with very high scores of 92% for the learners and 91% for the native speakers. Our L2 group's high accuracy scores are in stark contrast to those found in some earlier L2 processing studies that have used metalinguistic judgment tasks (Juffs & Harrington, 1996; Williams et al., 2001 – but see Juffs, 2004). The high degree of variation in L2 learners' offline responses to GP sentences across different studies could be indicative of strong task effects in L2 processing (compare also Williams, 2006), an issue that is clearly in need of further investigation.

Conclusion

Our findings show L2 learners are highly sensitive to plausibility information in their online sentence processing when reading for meaning, and this affects the process of recovery from syntactic misanalysis. While our learners showed evidence of online recovery from 'weak' garden paths, 'strong' GP sentences that required more substantial revisions to the current sentence representation were found to be more likely to lead to comprehension breakdown in L2 processing. The results of this study also suggest that it may be fruitful to take into account individual differences in terms of overall processing speed when investigating on-line sentence comprehension, in both L2 and native speaker studies. Clearly, however, more research is needed to investigate the role of plausibility information in L2 processing and how this interacts with other types of information, and to establish whether our findings generalize to other L1/L2 combinations.

References

Allan, D. (1992). *The Oxford Placement Test*. Oxford: Oxford University Press.

Baumann, H., Nagengast, J. & Klaas, G. (1993). 'New Experimental Setup (NESU)' ms. Max Planck Institute for Psycholinguistics, Nijmegen.

Centre for Lexical Information. (1993). *The Celex lexical database.* Nijmegen: Max Planck Institute for Psycholinguistics.

Christianson, K., Hollingworth, A., Halliwell, J. & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42,* 368-407.

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Misrepresentations of garden-path sentences by older and younger adults. *Discourse Processes, 42*(2), 205-238.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics,* 27, 3-42.

Clifton, C. (1993). Thematic roles in sentence parsing. *Canadian Journal of Experimental Psychology,* 47, 222-246.

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

Felser, C., Roberts, L, Gross, R., & Marinis, T. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, 24, 453-489.

Felser, C., Sato, M., & Bertenshaw, N. (2009). The on-line application of binding Principle A in English as a second language. *Bilingualism: Language and Cognition* 12(3), 485-502.

Ferreira F., & Henderson, J. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, 30, 725-745.

Fodor, J. D., & Ferreira, F. (1998). *Reanalysis in Sentence Processing.* Dordrecht, NL: Kluwer.

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

Frazier, L. & Clifton, C. (1998). Sentence reanalysis and visibility. In J.D. Fodor & F. Ferreira (Eds). *Reanalysis in Sentence Processing.* Dordrecht, NL:Kluwer.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*(2), 178-210.

Frenck-Mestre, C. (2005). Ambiguities and anomalies: What can eye-movements and event-related potentials reveal about second language sentence processing? In J. Kroll & A. De Groot (Eds.). *Handbook of Bilingualism.* Oxford: Oxford University Press.

Gibson, E. (1991). *A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown.* Unpublished Ph.D. dissertation, Carnegie Mellon.

Gorrell, P. (1995). *Syntax and Parsing.* Cambridge: Cambridge University Press.

Grodner, D., Gibson, E., Argaman, V., & Babyonyshev, M. (2003). Against repair-based reanalysis in sentence comprehension. *Journal of Psycholinguistic Research, 32*(2), 141-166.

Hahne, A., & Friederici, A. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition,* 4, 123-141.

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32(2), 101-124.

Havik, E., Roberts, L., van Hout, R., Schreuder, R. and Haverkort, M. (2009). Processing subject-object ambiguitities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning,* 59(1), 73-112.

Holmes, V. M., Kennedy, A., & Murray, W. (1987). Syntactic structure and the garden path. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology,* 39A, 277-294.

Holmes, V., Stowe, L., & Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language,* 28, 668-689.

Juffs, A. (2004). Representation, processing, and working memory in a second language. *Transactions of the Philological Society,* 102, 199-225.

Juffs, A., & Harrington, M. (1996). Garden path sentences and error data in second language processing research. *Language Learning,* 46, 286-324.

Just, M., Carpenter, P., & Woolley, J. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General,* 111, 228-238.

Kennedy, A., Murray, W., Jennings, F., & Reid, C. (1989). Parsing complements: Comments on the generality of the principle of minimal attachment. *Language and Cognitive Processes* 4, 51-76.

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

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition,* 27, 53-78.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the Influence of Thematic Fit (and Other Constraints) in On-line Sentence Comprehension. *Journal of Memory and Language, 38*(3), 283-312.

Papadopoulou, D. (2005). Reading-time studies of second language ambiguity resolution. *Second Language Research,* 21, 98-120.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, 24, 501-528.

Papadopoulou, D., & Tsimpli, I. (2005). Morphological cues in children's processing of ambiguous sentences: a study of subject/object ambiguities in Greek. In A. Brugos, M.-R. Clark-Cotton & S. Ha (Eds.), *BUCLD 29: Proceedings of the 29th Annual Boston University Conference on Language Development* (pp. 471-481). Somerville, MA: Cascadilla Press.

Pickering, M., & Traxler, M. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 24, 940-961.

Pickering, M., & Van Gompel, R. (2006). Syntactic parsing. In M. Traxler & M. Gernsbacher (Eds.). *The handbook of psycholinguistics* (pp. 455-503). San Diego, CA: Elsevier.

Pritchett, B. (1992). *Grammatical competence and parsing performance.* Chicago: University of Chicago Press.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin, 124*(3), 372-422.

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition,* 30, 333-357.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic Ambiguity Resolution in Discourse: Modeling the Effects of Referential Context and Lexical Frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*(6), 1521-1543.

Stowe, L. (1989). Thematic structures and sentence comprehension. In G. Carlson & M. Tanenhaus (Eds.). *Linguistic structure in language processing* (pp. 319-356). Dordrecht, NL: Kluwer.

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

Sturt, P., Pickering, M., & Crocker, M. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language,* 40, 136-150.

Trueswell, J. C., & Kim, A. E. (1998). How to Prune a Garden Path by Nipping It in the Bud: Fast Priming of Verb Argument Structure. *Journal of Memory and Language, 39*(1), 102-123.

van Gompel, R. P. G., & Pickering, M. J. (2001). Lexical guidance in sentence processing: A note on Adams, Clifton, and Mitchell (1998). *Psychonomic Bulletin & Review, 8*, 851-857.

Vasishth, S., & Lewis, R. L. (2006). Symbolic models of human sentence processing. In *Encyclopedia of Language and Linguistics, 2nd Edition.*: Elsevier.

Weinberg, A. (1999). A minimalist theory of human sentence processing. In S. Epstein & N. Hornstein (Eds.)., *Working minimalism* (pp. 283-315). Cambridge, MA: MIT Press.

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

Williams, J., Möbius, P., & Kim, C. (2001). Native and non-native processing of English *wh*-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics,* 22*,* 509-540.

Appendix

A. Experimental sentences and questions - 'Weak' GPs

1 The man believed the girl/book had upset very many people.

Had the girl/book upset many people?

2 The journalist wrote the book/girl had amazed all the judges.

Had the book/girl amazed all the judges?

3 The psychologist advised the lady/theory should be accepted by everyone.

Should the lady/theory be accepted by everyone?

4 The scientist proved the theory/lady could solve the difficult problem.

Could the theory/lady solve the problem?

5 The inspector warned the boss/crimes would destroy very many lives.

Would the inspector destroy many lives?

6 The criminal confessed the crimes/boss had been discovered too late.

Had the criminal been discovered too soon?

7 The dean warned the professor/magazine would never get published again.

Would the dean get published again?

8 The woman read the magazine/professor had shocked the university staff.

Had the woman shocked university staff?

9 The man warned the doctor/story would embarrass the hospital managers.

Would the doctor/story embarrass the hospital managers?

10 The spokesman confirmed the story/doctor had surprised the president yesterday.

Had the story/doctor surprised the president?

11 The student suggested the answer/teacher was not in fact correct.

Was the student not in fact correct?

12 The headmaster cautioned the teacher/answer would not be clearly understood.

Would the headmaster be clearly understood?

13 The manager resolved the issue/report would be discussed very soon.

Would the issue/report soon be discussed?

14 The student wrote the report/issue would start an important debate.

Would the report/issue start an important debate?

15 The headmaster cautioned the tutor/poem was unsuitable for the class.

Was the headmaster unsuitable for the class?

16 The man read the poem/tutor had excited the young children.

Had the poem/tutor bored the children?

17 The teacher cautioned the mother/problem would not go away easily.

Would the teacher go away easily?

18 The judge resolved the problem/mother would be described in court.

Would the judge be described in court?

19 The minister advised the king/truth should be discussed in parliament.

Should the king/truth be discussed in parliament?

20 The man confessed the truth/king was not believed by anyone.

Did the man make a confession?

B. Experimental sentences and questions - 'Strong' GPs

1	While the band played the song/beer pleased all the customers.	
	Did the song/beer please the customers?	
2	As the men drank the beer/song pleased everybody very much.	
	Did the beer/song make everybody unhappy?	
3	As the window-cleaner called housewife/ladder started to shake alarmingly.	
	Did the housewife/ladder start to shake?	
4	While the child climbed the ladder/housewife fell to the ground.	
	Did the child fall to the ground?	
5	While the neighbour visited the boy/car passed by the house.	
	Did the neighbour pass by the house?	
6	As the driver parked the car/boy made a loud noise.	
	Did the driver make a loud noise?	
7	While the woman drank the coffee/flowers looked very nice indeed.	
	Did the coffee/flowers look very nice?	
8	As the girl painted the flowers/coffee smelt very fresh indeed.	
	Did the flowers/coffee smell very fresh?	
9	As the woman called the puppy/ice-cream fell to the ground.	
	Did the puppy/ice-cream fall to the ground?	
10	While the child ate the ice-cream/puppy dropped to the floor.	
	Did the ice-cream/puppy drop to the floor?	
11	While the cleaner polished the truck/cake was stolen by someone.	
	Was the cleaner stolen by someone?	
12	As the mother baked the cake/truck disappeared mysteriously from sight.	
	Did the mother disappear from sight?	

13 As the woman ate the fish/tree shone in the sun.

Did the woman shine in the sun?

14 While the boy climbed the tree/fish looked very beautiful indeed.

Did the tree/fish look very ugly?

15 While the boy walked the dog/milk got hot and smelly.

Did the boy get hot and smelly?

16 As the girl drank the milk/dog disappeared from the kitchen.

Did the girl disappear from the kitchen?

17 As the woman rode the horse/boat raced under the bridge.

Did the horse/boat race under the bridge?

18 While the captain sailed the boat/horse passed by very quickly.

Did the boat/horse pass by quickly?

19 As the pilot flew the plane/song sounded far too loud.

Did the plane/song sound far too loud?

20 As the choir sang the song/plane could be heard by everyone.

Could the song/plane be heard by everyone?

Table 1. Summary of the L2 learners' bio-data and proficiency scores.

		Age[a]	Age of first exposure[a]	OPT scores[b]
Greek (*n*=25)	*Mean*	24.24	8.38	175.96
	Range	20-30	6-11	153-196
	SD	2.57	1.48	9.99

[a] In years.

[b] The means represent scores out of a possible maximum of 200.

Table 2. Accuracy scores % (SDs) for comprehension questions following the 'weak' GP items.

	Plausible DO	Implausible DO
Greek L2 Learners	95 (7)	90 (11)
Native Speakers	86 (10)	91 (08)

Table 3. Results of ANOVAs on the ambiguous noun (Segment 5) and the following five segments and on the summed reading time of the comprehension questions for 'weak' GPs.

		Segment 5 (boss/crimes)	Segment 6 (would)	Segment 7 (destroy)	Segment 8 (very)	Segment 9 (many)	Segment 10 (lives.)	Comp Q
Plausibility	*F1 (1, 47)*	4.25* (.08)	6.41* (.12)	2.75	4.94* (.10)	.467	2.31	5.51* (.11)
	F2 (1, 38)	3.63$^{(*)}$ (.09)	5.75* (.13)	8.44** (.16)	4.99* (.12)	.565	.973	2.94
Plausibility * Group	*F1 (1, 47)*	3.35$^{(*)}$ (.07)	1.30	1.22	4.05* (.08)	.001	3.34	1.10
	F2 (1, 38)	4.62* (.11)	1.59	3.35	1.93	.083	.334	0.75
Group	*F1 (1, 47)*	25.15** (.35)	17.82** (.28)	14.64** (.28)	17.73** (.27)	18.27** (.26)	23.21** (.33)	11.60** (.20)
	F2 (1, 38)	92.28** (.71)	62.46** (.62)	51.49** (.58)	45.84** (.57)	43.76** (.54)	43.18** (.79)	30.28** (.44)

$p < 0.08 = {}^{(*)}$

$p < 0.05 = *$

$p < 0.01 = **$

Partial η^2 reported in brackets for significant effects.

Table 4. Accuracy scores % (SDs) for comprehension questions following the experimental items in the 'strong' GP items.

	Plausible DO	Implausible DO
Greek L2 Learners	86 (13)	96 (6)
Native Speakers	95 (9)	93 (10)

Table 5. Results of ANOVAs on the ambiguous noun (Segment 6) and the following four segments, 'strong' GPs

		Segment 6 (song/beer)	Segment 7 (pleased)	Segment 8 (all)	Segment 9 (the)	Segment 10 (customers)	Comp Q
Plausibility	*F1 (1, 47)*	4.47* (.09)	3.26	3.03	1.26	.230	.751
	F2 (1, 38)	9.36** (.20)	2.23	3.33	3.29$^{(*)}$ (.08)	.350	.688
Plausibility * Group	*F1 (1, 47)*	4.20* (.08)	4.38* (.09)	.485	.005	.883	.209
	F2 (1, 38)	8.98* (.19)	3.29$^{(*)}$ (.08)	.688	.038	.460	.899
Group	*F1 (1, 47)*	20.75** (.31)	31.77** (.40)	34.35** (.42)	17.01** (.27)	31.30** (.40)	19.69** (.30)
	F2 (1, 38)	67.74** (.64)	75.29** (.67)	31.10** (.45)	43.89** (.54)	63.82** (.63)	33.28** (.47)

$p < 0.08 = {}^{(*)}$

$p < 0.05 = *$

$p < 0.01 = **$

Partial η^2 reported in brackets for significant effects

Table 6. Two speed groups: Native speakers' mean RTs for 'weak' GP items

	Seg_5 (boss/crimes)	Seg_6 (would)	Seg_7 (destroy)	Seg_8 (very)	Seg_9 (many)	Seg_10 (lives.)
FastNS-pl	231	221	247	229	168	304
FastNS-impl	212	227	259	240	173	335
SlowNS-pl	349	420	436	405	331	561
SlowNS-impl	378	477	398	388	307	544

Table 7. Native Speakers: Results of ANOVAs with Reading Speed as a covariate on the ambiguous noun (Segment 5) and the following five segments, 'weak' GPs

		Segment 5 (boss/crimes)	Segment 6 (would)	Segment 7 (destroy)	Segment 8 (very)	Segment 9 (many)	Segment 10 (lives.)	CompQ RTs	Accuracy
Plausibility	*F1 (1, 22)*	6.41* (.23)	1.910	2.20	3.73	.091	.376	.193	5.76* (.11)
Plausibility * Speed	*F1 (1, 22)*	7.57* (.27)	4.46* (.17)	3.26	4.32* (.16)	.283	.427	3.85	5.81* (.11)
Speed	*F1 (1, 22)*	91.01** (.80)	240.23** (.92)	93.41** (.81)	141.35** (.86)	88.61** (.80)	73.96** (.77)	80.78** (.79)	73.96** (.77)

$p < 0.08 = {}^{(*)}$

$p < 0.05 = *$

$p < 0.01 = **$

Partial η^2 reported in brackets for significant effects

Table 8. Two speed groups: Native speakers' mean RTs for 'strong' GP items

	Seg_6 (song/beer)	Seg_7 (pleased)	Seg_8 (all)	Seg_9 (the)	Seg_10 (customers.)
FastNS-pl	223	266	279	171	302
FastNS-impl	217	290	252	178	297
SlowNS-pl	409	559	443	327	543
SlowNS-impl	419	520	420	275	572

Table 9. Native Speakers: Results of ANOVAs with Reading Speed as a covariate on the ambiguous noun (Segment 6) and the following four segments, 'strong' GPs

		Segment 6 (song/beer)	Segment 7 (pleased)	Segment 8 (all)	Segment 9 (the)	Segment 10 (customers.)	CompQ RTs	Accuracy
Plausibility	*F1 (1, 22)*	.073	5.52* (.20)	.002	7.92* (.27)	.045	.063	.453
Plausibility * Speed	*F1 (1, 22)*	.053	6.54* (.20)	.188	12.30** (.36)	.182	.001	.940
Speed	*F1 (1, 22)*	133.81** (.86)	249.38** (.92)	94.72** (.81)	88.61** (.80)	75.50** (.77)	83.50** (.79)	1.84

$p < 0.08 = {}^{(*)}$

$p < 0.05 = *$

$p < 0.01 = **$

Partial η^2 reported in brackets for significant effects

Table 10. Two speed groups: L2 learners' mean RTs for 'weak' GP items

	Seg_5 (boss/crimes)	Seg_6 (would)	Seg_7 (destroy)	Seg_8 (very)	Seg_9 (many)	Seg_10 (lives.)
FastL2s-pl	424	398	412	374	331	697
FastL2s-impl	445	454	382	372	294	662
SlowL2s-pl	631	748	758	690	495	1150
SlowL2s-impl	789	860	655	573	515	1024

Table 11. L2 learners: Results of ANOVAs with Reading Speed as a covariate on the ambiguous noun (Segment 5) and the following five segments, 'weak' GPs

		Segment 5 (boss/crimes)	Segment 6 (would)	Segment 7 (destroy)	Segment 8 (very)	Segment 9 (many)	Segment 10 (lives.)	CompQ RTs	Accuracy
Plausibility	*F1 (1, 23)*	5.35* (.19)	122	.859	4.46* (.16)	3.11	1.05	.998	.394
Plausibility * Speed	*F1 (1, 23)*	9.51** (.29)	2.96	2.96	8.80** (.28)	2.97	2.52	.701	.697
Speed	*F1 (1, 23)*	72.54** (.76)	91.55** (.78)	91.55** (.78)	276.98** (.92)	53.45** (.70)	68.10** (.75)	50.24** (.69)	3.66

$p < 0.08 = {}^{(*)}$

$p < 0.05 = *$

$p < 0.01 = **$

Partial η^2 reported in brackets for significant effects

Table 12. Two speed groups: L2 learners' mean RTs for 'strong' GP items

	Seg_6 (song/beer)	Seg_7 (pleased)	Seg_8 (all)	Seg_9 (the)	Seg_10 (customers.)
FastL2s-pl	418	573	509	318	701
FastL2s-impl	414	660	429	287	684
SlowL2s-pl	690	925	652	529	1082
SlowL2s-impl	926	1041	616	518	1022

Table 13. L2 learners: Results of ANOVAs with Reading Speed as a covariate on the ambiguous noun (Segment 6) and the following four segments, ‘strong' GPs

		Segment 6 (song/beer)	Segment 7 (pleased)	Segment 8 (all)	Segment 9 (the)	Segment 10 (customers.)	CompQ RTs	Accuracy
Plausibility	*F1 (1, 23)*	3.91* (.15)	.271	587	4.55* (.17)	.193	.091	.210
Plausibility * Speed	*F1 (1, 23)*	7.51* (.25)	.569	840	4.12 (*) (.15)	.061	.032	.850
Speed	*F1 (1, 23)*	113.24** (.83)	50.91** (.69)	46.52** (.67)	53.29** (.70)	51.89** (.69)	61.75** (.73)	.752

$p < 0.08 = {}^{(*)}$

$p < 0.05 = *$

$p < 0.01 = **$

Partial η^2 reported in brackets for significant effects

Figure 1. Word-by-word reading times (in milliseconds) for complement clause items.

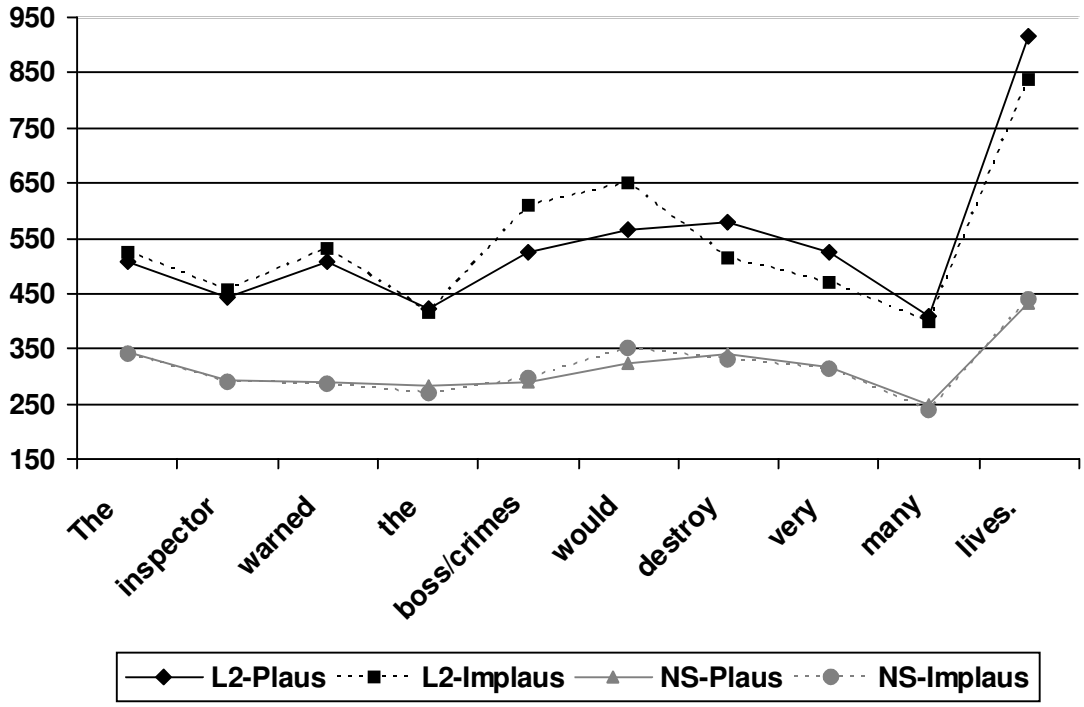
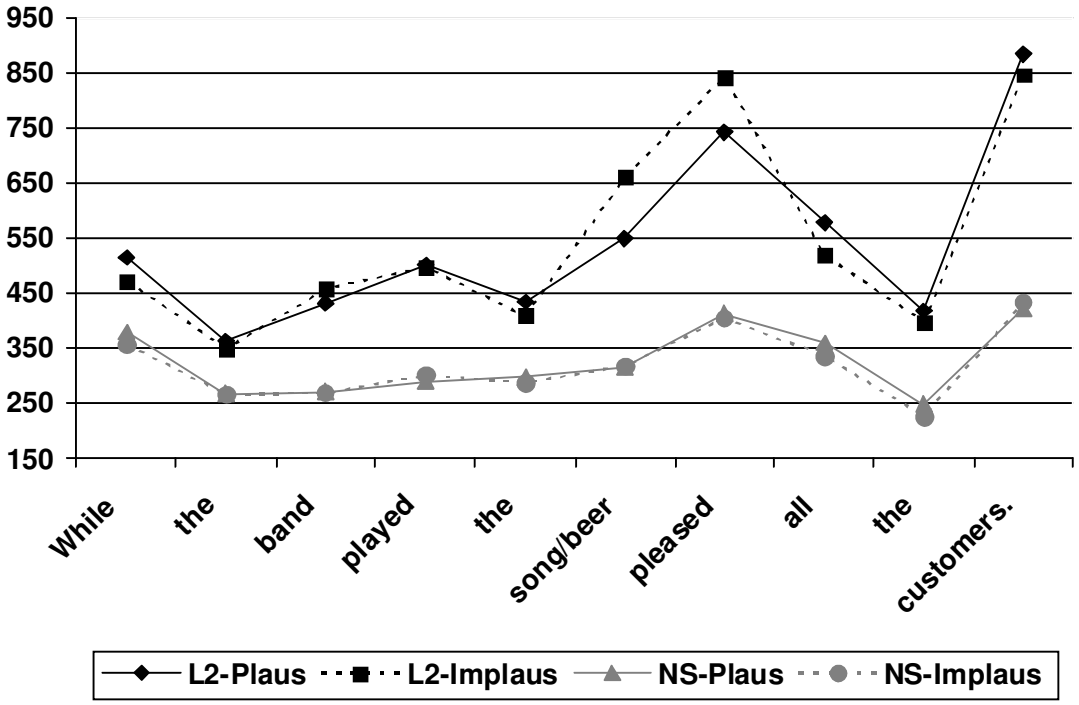


Figure 2. Word-by-word reading times (in milliseconds) for adjunct clause items.



Endnotes

[1] A reprocessing approach (e.g., Grodner, Gibson, Argaman, & Babyonyshev, 2003) also predicts this difference, but on the basis of locality differences: the ambiguous NP becomes the subject of a new clause that in (3) is connected locally to the main verb *saw*, whereas the new subject must be semantically and syntactically linked to the earlier clause via the (non-local) clausal connective *as*.

[2] Task effects on L2 learners' on-line processing have been observed, for instance, Havik, Roberts, van Hout, Schreuder, & Haverkort (2009) found that German L2 learners' online processing of subject-object ambiguities in Dutch were affected by whether they were required to make a truth-value judgment of the sentence immediately after reading it, or whether they merely had to read for comprehension.

[3] Greek translations of our complement clause items would not be ambiguous (an overt complementiser is required), however, very similar garden-path sentences to our adjunct clause items can be created in Greek (see e.g. Papadopoulou & Tsimpli, 2005).

[4] We thank Roger van Gompel (p.c.) for suggesting this.

[5] Our materials differed from those in Pickering and Traxler (1998) in that we did not have intervening padding material between the ambiguous NP and the disambiguating VP.

[6] We follow Pickering and Traxler (1998) and Williams et al. (2001) in assuming that the use of non-garden-path control conditions for comparison is problematic here, since they would differ structurally from the experimental items.

[7] A reviewer notes that the RTs go in the opposite direction, and therefore there may have been a speed-accuracy trade off.

[8] We thank the two reviewers for this suggestion.

[9] The L2 learners' mean reading speed was significantly negatively correlated with their proficiency scores in English ($r = -.48$, $n = 25$, $p < 0.02$) such that the higher the learner's

proficiency, the faster their reading speed (mean proficiency fast group = 179, slow group = 172).

[10] A reviewer points out that the results of the speed analysis support the idea that L2 learners are more strongly influenced by semantic information, and that this may be a side-effect of their slower reading in comparison to native speakers. Specifically, when reading slowly, there is more time for semantic information to be evaluated and put to use in interpretation.

[11] A reviewer notes that it is also possible that Pickering and Traxler (1998) observed on-line effects because they used eye-tracking, which is thought to be more sensitive a tool than self-paced reading.

[12] The question of whether or not the use of plausibility information is delayed in monolingual sentence processing is still controversial, however (see Pickering & Van Gompel, 2006, for review).

[13] We thank an anonymous reviewer for pointing this out.

[14] According to Weinberg (1999), the adjunct clause (including the putative direct object NP) in this type of GP sentence will undergo 'spellout' – that is, it will be deemed a phonologically and semantically complete subunit - because the disambiguating verb cannot be integrated without breaking the previous command path. Once a constituent or sentence chunk has undergone spellout, any subsequent changes to its internal structure may require conscious (and hence, measurable) processing effort.

[15] As pointed out by a reviewer, this was the likely reason why Pickering & Traxler (1998) added an additional 'padding' phrase between the noun and the disambiguating verb.