The effects of L2 proficiency level on the processing of *wh*-questions among Dutch second language speakers of English[1]

CARRIE N. JACKSON AND JANET G. VAN HELL

Abstract

Using a self-paced reading task, the present study explores how Dutch-English L2 speakers parse English wh-*subject-extractions and* wh-*object-extractions. Results suggest that English native speakers and highly-proficient Dutch–English L2 speakers do not always exhibit measurable signs of on-line re-analysis when reading subject- versus object-extractions in English. However, less-proficient Dutch-English L2 speakers exhibit greater processing costs on subject-extractions relative to object-extractions, similar to previously reported findings (e.g., Dussias and Piñar 2010; Juffs 2005; Juffs and Harrington 1995). These findings are discussed in light of relevant research surrounding on-line processing among L2 speakers and their ability to adopt native-like processing patterns in the L2.*

1. This study was funded by a grant from the *Center for Language Science* at Pennsylvania State University. This study was also funded in part by grant R03HD058765 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development to Carrie N. Jackson and NSF Grant OISE-0968369 and NSF Grant BCS-0955090 awarded to Janet van Hell. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health & Human Development or the National Institutes of Health. We would like to thank Petra Aast, Tom Damen, Sjaron Jenneskens, Renske Leenders, Judith Siers, Ninke van Timmeren, and Marijt Witteman, for their help in collecting the Dutch-English L2 speaker data and Giulia Borriello, Kinsey Hallinger, Michael Lipschultz, Matthew Leaman, Emily Simms, and Stephanie Yessler for their help in collecting the English native speaker data. We would also like to thank Jared Linck and CJ Powers for their assistance with the statistical analyses. An earlier version of this paper was presented at the Center for Language Science at Penn State. We are grateful to the attendees for their insightful comments. Finally, we would like to thank the anonymous *IRAL* reviewers for their comments on earlier versions of this article. All errors are, of course, our own.

IRAL 49 (2011), 195–219
DOI 10.1515/iral.2011.012

0019042X/2011/049-195

1. Introduction

While there is a longer tradition in psycholinguistics of examining how second language (L2) speakers access words in each of their languages (see Dijkstra 2005; Kroll and Tokowicz 2005, for recent reviews), more recently researchers have begun to investigate how L2 speakers process lexical and morphosyntactic information during the on-line comprehension of L1 and L2 sentences (see Clahsen and Felser 2006; van Hell and Tokowicz 2010, for recent reviews). Much of this research has focused on whether L2 speakers apply potentially universal strategies to the parsing of L2 sentences and whether such strategies parallel those used by native speakers (e.g., Felser et al. 2003; Havik et al. 2009; Juffs 2005; Marinis et al. 2005; Williams 2006). At their core, many of these studies ask whether L2 speakers can ever achieve native-like abilities in the L2, or if there are inherent constraints in adult L2 acquisition.

To date, a majority of L2 sentence processing research has focused on highly proficient L2 speakers immersed in an L2 environment at the time of testing. Studies that have investigated on-line processing among less-proficient L2 populations, especially those whose exposure to the target language has been largely through classroom-based instruction, have looked primarily at learners' sensitivity to particular L2 grammatical information (e.g., Keating 2009; Osterhout et al. 2006; Tokowicz and MacWhinney 2005), as opposed to examining their on-line parsing preferences or how they build the grammatical structure of a sentence in real time (but see Frenck-Mestre 1997; Dekydtspotter et al. 2008; Jackson 2008, for exceptions to this trend). However, to build adequate models of how L2 speakers process L2 input, L2 proficiency is an important variable to consider. The present study addresses this issue by investigating how both highly-proficient and intermediate Dutch-English L2 speakers process English *wh*-questions, while also testing English native speakers to provide a baseline for comparison.

1.1. *Proficiency and L2 on-line processing*

While a growing body of research has investigated on-line processing among highly- proficient L2 speakers, fewer studies have explored how less-proficient L2 speakers piece together the meaning and structure of a sentence in real time. In one study that has focused on less-proficient L2 speakers, Frenck-Mestre (1997) compared English-French and Spanish-French L2 speakers who had been living in France for nine months at the time of testing, investigating how they processed temporarily ambiguous sentences, like (1), in their L2 French.

(1) *Jean connaît les filles de la gardienne qui partent . . .*
John knows the girls-PL of the nanny-SG who leave-3PL
'John knows the girls of the nanny who are leaving . . . ' (Frenck-Mestre 1997: 475)

Frenck-Mestre found that it was easier for English-French L2 speakers to interpret sentences in which subject-verb agreement information in the relative clause forced one to interpret the relative clause as providing additional information about the second noun of the complex noun phrase (i.e., *la gardienne* 'the nanny'), often referred to as low attachment. This pattern parallels a preference for low attachment previously found among monolingual English speakers (e.g., Frazier and Clifton 1996). In contrast, both Spanish-French L2 speakers and French native speakers found it easier to interpret the relative clause as describing the first noun of the complex noun phrase (i.e., *les filles* 'the girls'), often referred to as high attachment. This pattern parallels a preference for high attachment previously found among native speakers of both French and Spanish (e.g., Cuetos and Mitchell 1988; Zagar et al. 1997). Given that the reading patterns for the L2 groups differed from each other in their L2 French, but mirrored each group's respective L1 preferences, Frenck-Mestre argued that at least at lower proficiency levels, L2 speakers may transfer on-line processing preferences from their L1 when reading L2 sentences. Interestingly, in a later study, Frenck-Mestre (2002) reported that English-French and Spanish-French L2 speakers who had been living in France for five years at the time of testing exhibited a native-like French preference for high attachment, suggesting that with increased proficiency, L2 speakers can exhibit L2 on-line processing preferences that parallel those of native speakers.

Hopp (2006) also found that L2 proficiency played a critical role in L2 speakers' reading time patterns when processing subject- and object-first sentences in the L2 German.

(2) *Er denkt, dass den Physiker am Freitag* (object-first)
he thinks, that the$_{ACC}$ physicist on Friday
der Chemiker gegrüsst hat.
the$_{NOM}$ chemist greeted has
'He thinks that the chemist greeted the physicist on Friday.'

(3) *Er denkt, dass der Physiker am Freitag* (subject-first)
he thinks, that the$_{NOM}$ physicist on Friday
den Chemiker gegrüsst hat.
the $_{ACC}$ chemist greeted has
'He thinks that the physicist greeted the chemist on Friday.' (Hopp 2006: 378)

Near-native English- and Dutch-German L2 speakers took longer to read object-first sentences, like (2), than subject-first sentences, like (3), immediately at the disambiguating noun phrase (underlined above), similar to German native speakers. However, advanced – but not near-native – English- and Dutch-German L2 speakers only exhibited longer reading times on object-first sentences compared to subject-first sentences at the final word in the sentence (e.g., *hat* 'has'), suggesting that the advanced L2 speakers were not able to process disambiguating case-marking information immediately at the disambiguating region. However, similar to Frenck-Mestre's studies (1997, 2002), even the less-proficient L2 speakers in Hopp's study were living in an L2-dominant environment at the time of testing and were, arguably, more proficient than L2 speakers whose exposure to the L2 has occurred primarily in a foreign language classroom. Thus, these studies still leave open the question of how classroom-based learners process L2 sentences in real time.

More recently, Dekydtspotter et al. (2008) conducted a study that focused precisely on classroom-based learners, examining whether English-French L2 speakers in their second or forth semesters of L2 learning showed on-line relative clause attachment preferences in their L2 French. Participants completed a self-paced reading task in which the preferred attachment site for the relative clause was biased towards either the first noun (i.e., high attachment) or the second noun (i.e., low attachment) via contextual information presented prior to the target sentence. Dekydtspotter et al. reported that the L2 speakers took longer to read the verb in the relative clause in sentences in which contextual information forced participants to attach the relative clause to the first noun than in sentences in which contextual information supported a low attachment interpretation. Dekydtspotter et al. emphasized that their findings demonstrate that even less-proficient L2 speakers have recourse to syntactic information during on-line processing. Specifically, had the L2 speakers relied exclusively on contextual information to interpret the target sentences, there should have been no increase in reading times on the conditions in which the biasing contextual information forced participants to compute a dispreferred syntactic structure, as was the case in high-attachment sentences. Instead, even the second-semester L2 speakers exhibited longer reading times at the relative clause verb for high-attachment contexts relative to low-attachment contexts. Thus, even at lower proficiency levels, Dekydtspotter et al. argue that L2 speakers are sensitive to structural constraints during L2 on-line processing, leading to increased processing costs when the required syntactic structure of a sentence imposes a greater processing load. Such findings are particularly relevant as L2 processing research debates the extent to which L2 speakers are able, or unable, to construct a detailed syntactic representation of a sentence in real time (see Clahsen and Felser 2006).

At the same time, the target sentences employed by Dekydtspotter et al. (2008) were not disambiguated via morphosyntactic information in the target sentence itself (e.g., subject-verb agreement), rather via pragmatic biases in the context presented prior to the target sentence. Therefore, it is possible that unintended confounds in the preceding context, such as the number of times a particular referent was mentioned, may have influenced the overall plausibility of the preceding context in the first place, leading to reading time differences on the final target sentence independent of the high- vs. low-attachment manipulation. This possibility is supported by the fact that both the second and the fourth semester participants responded at close to chance levels to an accompanying question that asked whether the target sentence was an acceptable continuation of the biasing context, regardless of whether the context was supposed to bias one towards a high- or low-attachment interpretation. Thus, additional work is needed to substantiate Dekydtspotter et al.'s claim that even less-proficient L2 speakers exhibit immediate sensitivity to structural information during real time language processing in their L2.

1.2. *L2 processing of filler-gap constructions*

Related to the debate surrounding L2 speakers' ability to use structurally-based parsing strategies in real time, several studies have examined how L1 and L2 speakers process so-called filler-gap constructions, such as *Who did the boy believe he saw on the playground?*. Although different theories exist regarding the exact nature of how the human parsing mechanism deals with such constructions, converging cross-linguistic evidence suggests that both monolingual and highly-proficient L2 speakers will attempt to integrate an initial *wh*-phrase into the sentence as soon as possible (Dussias and Piñar 2010; Jackson and Dussias 2009; Juffs and Harrington 1995; Williams 2006; Williams et al. 2001; for various theoretical accounts for this phenomenon see Carlson and Tanenhaus 1988; De Vincenzi 2000; Fodor 1993; Frazier 1987; Gibson 1998; Pritchett 1992). For example, Juffs and Harrington (1995) and Dussias and Piñar (2010) found that Chinese-English L2 speakers exhibited longer reading times at the complement clause (i.e., *killed the pedestrian*) on subject-extractions, like (4), compared to object-extractions, like (5).

(4) *Who did the police know killed the pedestrian?* (subject-extraction)

(5) *Who did the police know the pedestrian killed?* (object-extraction)
(Dussias and Piñar 2010)

Juffs and Harrington, as well as Dussias and Piñar, proposed that both native and L2 participants used a gap-filling parsing strategy in which they attempted to integrate the initial *wh*-phrase, *who*, into the first clause upon reading the

verb *know* (i.e., *Who did the police know*), thereby initially assigning the thematic role of direct object to *who*. When participants subsequently read the complement clause (underlined in examples (4) and (5)), this initial role assignment became untenable in subject-extractions, leading to increased processing costs for subject-extractions relative to object-extractions. Dussias and Piñar also found a similar processing asymmetry at the complement clause among English native speakers. In contrast, Juffs and Harrington found no parallel processing asymmetry in the English native speakers in their study, leading them to argue that although highly-proficient L2 speakers and English native speakers may process such *wh*-questions in a similar manner, this may not always lead to measurable signs of on-line reanalysis among native speakers as compared to L2 speakers of English.

While not testing this specific type of *wh*-question, there is additional evidence from the monolingual sentence processing literature to suggest that reanalysis during the on-line processing of filler-gap constructions need not always be costly. For instance, Kaan (1997) and Bader and Meng (1999) have reported that native speakers of Dutch and German exhibit fewer processing difficulties when reading temporarily ambiguous *wh*-questions compared to other types of filler-gap constructions, like relative clauses.[2] Dussias and Piñar (2010) also reported that English native speakers and high-working memory Chinese-English L2 speakers had greater difficulty processing subject-*wh*-extractions relative to object-*wh*-extractions, even in conditions in which the initial *wh*-phrase was implausible as the direct object of the main verb (e.g., *Who do you think* . . .). Thus, it is unlikely that the processing asymmetries reported by Juffs and Harrington and Dussias and Piñar can be attributed solely to difficulties in assigning or re-assigning thematic roles between subject- and object-extractions in their target sentences.

In a follow-up study, Juffs (2005) also reported processing costs for subject-extractions relative to object-extractions for Spanish- and Japanese-English L2 speakers, replicating other findings with L2 speakers.[3] However, in the same

2. Kaan (1997) and Bader and Meng (1999) also found that for both *wh*-questions and relative clauses, German and Dutch native speakers actually exhibit greater processing difficulties on object-first sentences than subject-first sentences (see also Frazier 1987; Frazier and Flores d' Arcais 1989). A similar subject-first preference has been reported for subject- vs. object-relative clauses in English (e.g., Traxler et al. 2002). This broader subject-first preference is the exact opposite of the processing asymmetry reported by Dussias and Piñar (2010) and Juffs and Harrington (1995) for subject- vs. object-*wh*-extractions in English. This provides additional evidence that how native and L2 speakers process English subject vs. object-*wh*-extractions may differ from other types of filler-gap constructions (but see Lee 2010 for counterevidence).
3. Juffs (2005) did not make any statistical comparisons between subject- and object-extractions for the English native speakers. Thus, it is unknown whether the English native speakers in his study exhibited increased processing costs on subject-extractions relative to object-extractions.

study he found no increase in reading times for L2 speakers when they read subject-extractions out of nonfinite clauses, such as (6).

(6) *Who does the boss expect ____ to meet the customers next Monday?* (subject-extraction; nonfinite clause)

(7) *Who does the boss expect the customers to meet ____ next Monday?* (object-extraction; nonfinite clause) (Juffs 2005: 129)

This led Juffs to hypothesize that the difficulty with subject-extractions out of finite clauses, like (4), may not stem from the difficulty of reassigning thematic roles to the initial *wh*-phrase but rather that such difficulties lie with processing two adjacent finite verbs. In subject-extractions out of finite clauses, the main verb is immediately followed by the verb of the complement clause, leading to longer reading times for subject-extractions out of finite clauses at the complement clause relative to either subject- or object-extractions out of nonfinite clauses, like (6) and (7), or object-extractions out of finite clauses, like (5). In spite of positing different explanations for the costs associated with processing subject-extractions in English, results from all three studies – Juffs and Harrington (1995), Dussias and Piñar (2010) and Juffs (2005) – suggest that L2 speakers adopt some sort of gap-filling strategy when processing *wh*-questions in the L2, regardless of their L1 background (see also Williams 2006; Williams et al. 2001).

However, Clahsen and Felser (2006) have correctly pointed out that the filler-gap position coincided with the main verb in the *wh*-questions used in the English *wh*-question studies summarized here (Dussias and Piñar 2010; Juffs 2005; Juffs and Harrington 1995; see also Williams 2006; Williams et al. 2001). Therefore, it is difficult to determine whether the processing difficulties encountered by the L2 participants in these studies stemmed from difficulties integrating the *wh*-phrase with its subcategorizing verb or whether the L2 participants actually employed a structure-based gap-filling strategy.

Further, other studies, such as Marinis et al. (2005) and Roberts and Felser (2007), have shown that L2 speakers may not pose an intermediate gap (indicated by dashes below) when reading long distance *wh*-extractions that extend over clause boundaries, like *The nurse who the doctor argued – that the rude patient had angered – is refusing to work late*. Specifically, reading times were longer for both English native speakers and L2 English speakers at *had angered* relative to non-extraction control sentences, consistent with increased processing costs at the point one must integrate a *wh*-phrase with its subcategorizing verb. However, this difficulty was reduced in sentences containing an intermediate landing site for the *wh*-phrase (e.g., at *argued*) for English native speakers, but there was no corresponding reduction in reading times among the L2 speakers in such conditions. Such findings led Clahsen and Felser (2006)

to suggest that grammatical processing among L2 speakers may be shallower and involve less detailed syntactic representations compared to that of native speakers (but see Dekydtspotter et al. 2006 for counterarguments).

To conclude, it remains unresolved whether highly proficient L2 speakers employ a structure-based or verb-driven gap filling strategy during on-line L2 processing.[4] However, there is converging evidence that at the very least, highly-proficient L2 speakers search for a landing site for an initial *wh*-phrase when processing *wh*-questions in their L2 and may exhibit processing difficulties at the point they must integrate a *wh*-phrase with its subcategorizing verb (Dussias and Piñar 2010; Felser and Roberts 2007; Jackson and Dussias 2009; Juffs 2005; Juffs and Harrington 1995; Marinis et al. 2005; Williams 2006; Williams et al. 2001). Up to now, however, the studies that have investigated how L1 and L2 speakers process *wh*-questions in English have tested highly-proficient L2 speakers living in an English immersion environment. Whether less-proficient L2 speakers of English – in particular those who have learned English primarily in a classroom setting – will also exhibit similar difficulties when reading L2 sentences remains an open question. Given the number of L1 and L2 sentence processing studies that have attempted to account for how native and L2 speakers parse *wh*-questions in English, this is an important question to consider, especially in view of other recent findings suggesting that even less-proficient L2 speakers may have recourse to structurally-based information when reading L2 sentences (e.g., Dekydtspotter et al. 2008).

2. Method

2.1. *Participants*

Thirty three advanced Dutch-English L2 speakers (ages: 18–34) were recruited from a large Dutch university and the surrounding community. Twenty five intermediate Dutch-English L2 speakers (ages: 15–17) were recruited from high schools in the same area. In accordance with the Dutch school system, all participants began learning English in 5th grade (age: 10–11). Basic L2 proficiency information for the Dutch-English L2 speakers is presented in Table 1.

As a preliminary indication of L2 proficiency, participants were asked to self-rate their L2 English proficiency on a 7-point scale (1 = do not speak the language, 7 = same as native language). T-tests compared the two L2 speakers'

4. As the stimuli in the current experiment were based on the same constructions as those used by Dussias and Piñar (2010), they were not designed to tease apart whether L2 speakers rely on a structure-based or verb-driven strategy when processing *wh*-dependencies in their L2. Therefore, we will not discuss this distinction further.

Table 1. *L2 proficiency information for Dutch–English L2 speakers (standard deviations in parentheses)*

	Intermediate L2 speakers (high school)			Advanced L2 speakers (university)		
	M		Range	M		Range
Self-ratings of L2 proficiency						
Reading	5.4	(1.1)	3–7	6.4	(0.5)	6–7
Writing	5.3	(1.0)	3–7	5.8	(0.8)	4–7
Listening	5.5	(1.1)	3–7	6.3	(0.6)	5–7
Speaking	5.3	(1.0)	3–7	5.8	(0.6)	5–7
English Llex proficiency score[a]	67.2	(13.7)	39–96	85.4	(6.9)	64–97

a Proficiency task score is out of 100%.

ratings according to group. The results of these analyses showed that the advanced Dutch-English L2 speakers considered themselves to be more proficient in English compared to the intermediate Dutch-English L2 speakers, although in the case of L2 writing skills, this difference only approached significance (all ps $< .05$; L2 writing: $t(56) = 1.89$, $p = .064$).

However, self-ratings are not necessarily a reliable measure of L2 proficiency because participants may under- or overestimate their actual proficiency in the language. Therefore, participants also completed the Llex English lexical decision task (Meara 1994) as an objective measure of L2 English proficiency. This task is a standardized lexical decision task containing real English words and pronounceable English pseudowords and participants must decide which letter strings are real English words. This test, or variations of this test, have been used as an independent measure of L2 proficiency in a number of previous L2 studies (e.g., Harley and Hart 1997; Hermans et al. 1998; Mady 2007; von Studnitz and Green 2002). While this test only explicitly tests L2 lexical knowledge, results from this task have also been shown to correlate with listening and reading comprehension in English (e.g., Harley and Hart 1997) and overall L2 English proficiency (Meara and Buxton 1987). The difference in mean scores on this lexical decision task between the intermediate and advanced L2 participants was significant ($t(56) = 6.62$, $p < .001$), further supporting the characterization of these two groups of L2 speakers as distinct with regard to their L2 proficiency.

Additionally, 24 English native speakers were recruited from a large public university in the United States (ages: 18–25). While the English native speaker participants reported some exposure to a foreign language, none considered

themselves fluent in anything other than English and none had spent any time immersed in an L2 environment.

2.2. *Materials*

The experiment consisted of 16 target stimuli and 64 filler items. The target stimuli were *wh*-questions in which the initial *wh*-phrase was either the subject or direct object of the complement clause, as in examples (8) and (9).

(8) *Who do you think* ____ *met the tourists in front of the museum?* (subject-extraction)

(9) *Who do you think the tourists met* ____ *in front of the museum?* (object-extraction)

Whether the initial *wh*-phrase was the subject or the direct object of the complement clause was disambiguated via word order in the critical region (underlined above). Each target sentence also included a prepositional phrase after the critical region, so that the critical region did not coincide with the end of the sentence. In the initial clause (e.g, *Who do you think* ...), we relied on three main verbs, namely *think, say*, and *suspect*. These three verbs were chosen from the larger set of verbs used by Dussias and Piñar (2010), as well as Juffs and Harrington (1995), and were chosen because they are all licit verbs for similar constructions in Dutch.[5]

With regard to the initial *wh*-phrase, although English still makes the distinction between *who* and *whom*, with *who* indicating the grammatical subject of a clause and *whom* indicating the direct object, whether English native speakers are sensitive to this distinction in either production or comprehension is open to debate (Aarts 1994). Furthermore, Dussias and Piñar (2010) found no difference in grammaticality judgment accuracy for similar subject- and object-extractions among either English native speakers or Chinese-English L2 speakers (see also Juffs 2005; Juffs and Harrington 1995), suggesting that neither group made a distinction between *who* and *whom* when judging the target sentences. Therefore, similar to these previous studies, we used *who* as the initial *wh*-phrase across experimental conditions, regardless of whether *who* ultimately was disambiguated as the subject or the direct object of the complement clause.

5. Pilot testing revealed that, contrary to findings reported by Dussias and Piñar (2010), using a wider range of main clause verbs and explicitly manipulating whether the initial *wh*-phrase *who* was a plausible direct object of the main verb (e.g. *Who do you think* ... vs. *Who do you know* ...) did not influence English native speakers' reading times at the critical complement clause. Specifically, English native speakers' reading times at the critical region and on the subsequent word did not vary as a function of either extraction type or plausibility.

To ensure that the target items were not biased towards a subject- or object-interpretation based on semantic information, 16 English native speakers and 16 advanced Dutch-English L2 speakers from the same university populations as the main experiment rated the plausibility of the target sentences on a 4-point scale, with 1 being "very plausible" and 4 being "very implausible". None of these participants participated in any other portion of the experiment. Among the English native speakers, the average plausibility rating for subject-extractions was 1.72 and the average plausibility rating for object-extractions was 1.79, a difference that was not statistically significant ($t(15) = .69$, $p = .504$), suggesting that they considered both sentence types equally plausible. The mean ratings from the Dutch-English L2 speakers were lower overall compared to the English native speakers (subject-extractions: $M = 2.59$; object- extractions: $M = 2.48$), indicating that they found the sentences less plausible overall than the English native speakers. However, there was no significant difference in their ratings for subject-extractions compared to object-extractions ($t(15) = .77$, $p = .451$). As a final precaution against potential bias effects, any sentence pair in which the mean ratings from either the English native speakers or the Dutch-English L2 speakers differed by more than 0.5 between the subject- and object-extraction conditions were modified and judged as being equally plausible across both conditions by an additional set of English native speakers.

In addition to the 16 target sentences, participants read 24 grammatical filler items and 40 ungrammatical filler items. These 64 filler items included additional grammatical and ungrammatical *wh*-questions, as well as grammatical and ungrammatical items that were designed to compare and contrast core features of Dutch and English grammar, such as word order differences (see Appendix for sample filler items). Thus, over the entire experiment there were an equal number of grammatical and ungrammatical sentences. The 16 target stimuli were divided into two lists, so that participants read 8 subject-extractions and 8 object-extractions, but they did not read more than one version of each target sentence. These 16 target stimuli were presented in a semi-randomized order along with the 24 grammatical filler items and the 40 ungrammatical filler items.

2.3. *Procedure*

Participants completed the task individually in a quiet room. The target and filler items were presented using the non-cumulative self-paced-reading paradigm (Just et al. 1982) via E-Prime experimental software (Schneider et al. 2002). Prior to beginning the experiment, participants were told that for each sentence they would see a row of dashes on the computer screen that repre-

sented each word in the sentence. Each time they pressed the space bar, the next word in the sentence would appear and the previous word would disappear. Before they began each sentence, the word READY appeared on the computer screen to indicate that participants should prepare themselves for the next sentence. After they pressed the space bar, this word disappeared and the first word of the sentence appeared.

Similar to previous studies examining this type of *wh*-question, when participants finished reading each sentence, a prompt appeared on the screen that asked them to judge whether the sentence they had just read was grammatical or ungrammatical. They responded by pressing Y for "yes" and N for "no" on a button box. It was decided to rely on grammaticality judgments rather than explicitly testing participants' comprehension of the target sentences to maximize the similarity between the present study and previous L2 sentence processing studies that have examined this particular syntactic structure (Dussias and Piñar 2010; Jackson and Dussias 2009; Juffs 2005; Juffs and Harrington 1995). Prior to beginning the experiment, participants were told orally that they should judge each sentence based on how it sounded, and that they should not rely on prescriptive grammar rules. However, no specific examples were given, so as not to bias participants. The participants completed 10 practice items at the beginning of the experiment to familiarize them with the self-paced reading format.

3. Results

3.1. *Judgment Accuracy*

Overall accuracy judgment rates for the target and filler items are presented in Table 2. A repeated-measures ANOVA was conducted comparing participants' judgment accuracy on the target stimuli, treating Extraction Type (subject-extraction vs. object-extraction) as a within-participants variable and Group (intermediate L2 speakers vs. advanced L2 speakers vs. English native speakers) as a between-participants variable in the participants analysis and a within-participant variable in the items analysis. This ANOVA revealed a main effect for Group ($F1(2,79) = 18.21$, $p < .001$; $F2(2,30) = 32.48$, $p < .001$). Follow-up t-tests, using the Bonferroni correction for multiple comparisons, compared the participant groups to each other. These t-tests revealed that the intermediate L2 speakers were less accurate in correctly judging the target stimuli as grammatical compared to the advanced L2 speakers and the English native speakers (intermediate vs. advanced: $t1(1,56) = 4.26$, $p < .001$; $t2(1,15) = 5.56$, $p < .001$; intermediate vs. native speakers: $t1(1,47) = 5.14$, $p < .001$; $t2(1,15) = 7.22$, $p < .001$), but there was no significant difference in

Table 2. *Grammaticality judgment results*

	Intermediate L2 speakers		Advanced L2 speakers		English native speakers	
	M	SD	M	SD	M	SD
Experimental items (overall)	79.3	14.5	92.2	8.6	95.6	5.7
Subject-extractions	76.0	24.2	90.5	12.9	96.4	5.8
Object-extractions	82.5	20.7	93.9	10.4	94.8	9.0
Filler items (overall)	72.1	12.9	90.1	9.1	93.5	4.5
Grammatical fillers	76.8	12.8	92.7	9.0	94.1	4.7
Ungrammatical fillers	69.0	17.3	89.4	12.4	93.1	6.6

judgment accuracy on the target stimuli between the advanced L2 speakers and the English native speakers, although the difference approached significance in the items analysis ($t1(1,55) = 1.66$, $p = .102$; $t2(1,15) = 1.89$, $p = .078$). There was no main effect for Extraction Type ($F1(1,79) = 1.26$, $p = .265$; $F2(2,30) = 2.15$, $p = .163$) and the interaction between Extraction Type and Group was not significant in the participants analysis but approached significance in the items analysis ($F1 < 1$; $F2(2,30) = 2.85$, $p = .074$).

In addition to the analysis of the target stimuli, two one-way ANOVAs were conducted comparing the judgment accuracy for the grammatical filler items and the ungrammatical filler items across the three participant groups. The results of each ANOVA revealed a significant effect for Group (grammatical filler items: $F(2,79) = 26.35$, $p < .001$; ungrammatical filler items: $F(2,79) = 25.74$, $p < .001$).[6] As with the results from the target stimuli, the intermediate L2 speakers' judgment accuracy on both the grammatical and ungrammatical filler items was significantly lower than the advanced L2 speakers' and the English native speakers' judgment accuracy (intermediate vs. advanced – grammatical: $t(56) = 5.52$, $p < .001$, ungrammatical: $t(56) = 5.22$, $p < .001$; intermediate vs. native speaker – grammatical: $t(47) = 6.19$, $p < .001$, ungrammatical: $t(47) = 6.37$, $p < .001$) but there were no significant differences in judgment accuracy for either the grammatical or ungrammatical filler items between the advanced L2 speakers and the English native speakers (grammatical: $t(55) = .703$, $p = .485$; ungrammatical: $t(55) = 1.33$, $p = .188$).

6. Two ungrammatical filler items were eliminated from these analyses because less than half of the English native speakers correctly judged these two sentences to be ungrammatical. Thus, the grammaticality judgment results for the ungrammatical fillers are based on 38 items.

Taken together, these results underscore the high level of English proficiency among the advanced L2 speakers, as their judgment accuracy did not differ significantly from the English native speakers on either the experimental or the filler items. At the same time, all participants from all three groups correctly judged at least 60 % of the target stimuli as accurate, and even the intermediate L2 speakers' mean judgment accuracy rates on the target sentences were similar to the grammaticality judgment accuracy rates for the L2 speakers reported by Juffs and Harrington (1995), Juffs (2005) and Dussias and Piñar (2010). This indicates that regardless of proficiency level, the participants understood the nature of the task and demonstrated a minimum level of grammatical awareness regarding the target *wh*-questions.

3.2. *Reading times*

Because subject- and object-extractions are disambiguated by word order in English, the critical region for the reading time analyses was defined as the verb and noun phrase in the complement clause (e.g., *met the tourists*). Reading times on these three words were averaged to obtain the reading time for the critical region (as in Dussias and Piñar 2010). The subsequent word was also analyzed to capture any potential spillover effects from processing the critical region. As is common in psycholinguistic literature, only results from items correctly judged as grammatical were included in the reading time analyses. Excluding data from incorrect responses resulted in the exclusion of 20.8 % of the data from the intermediate L2 speakers, 7.8 % of the data from the advanced L2 speakers and 4.4 % of the data from the English native speakers. Furthermore, all reading times less than 100 ms and greater than 6000 ms were excluded from analysis (10 cases) and for each condition within each participant group, reading times greater than 2.5 standard deviations were excluded. These data trims led to the exclusion of an additional 2.7 % of the intermediate L2 speaker data, 3.1 % of the advanced L2 speaker data and 2.1 % of the English native speaker data. Mean reading times for the critical region, as well as earlier words in the sentence, are presented in Table 3.

The reading time data were analyzed using hierarchical linear modeling (HLM). Advantages of HLM over more traditional ANOVA analyses include the ability to account for both participant and item variance simultaneously, eliminating the need for separate by-participant and by-item analyses (Baayen et al. 2008; Quené and van den Bergh 2008). Additionally, HLM does not require one to aggregate reading times across trials, nor does it require an equal number of observations for each individual (Baayen et al. 2008; Quené and van den Bergh 2008). These advantages are magnified when dealing with L2 data where lower accuracy rates may lead to the exclusion of a greater proportion of reading time data than comparable monolingual studies.

Table 3. *Mean reading times in milliseconds (standard deviations in parentheses)*

	Intermediate L2 speakers		Advanced L2 speakers		English native speakers	
Word 1 (*who*)						
Subject-extraction	573	(207)	452	(163)	505	(105)
Object-extraction	580	(180)	439	(129)	510	(125)
Word 2 (*do*)						
Subject-extraction	478	(102)	359	(104)	400	(86)
Object-extraction	495	(122)	363	(107)	397	(78)
Word 3 (*you*)						
Subject-extraction	482	(123)	355	(95)	399	(97)
Object-extraction	498	(116)	359	(90)	388	(69)
Word 4 (*think*)						
Subject-extraction	499	(131)	371	(104)	424	(97)
Object-extraction	494	(129)	388	(113)	412	(84)
Critical region (*met the tourists*)						
Subject-extraction	572	(178)	428	(134)	438	(98)
Object-extraction	529	(141)	426	(141)	444	(102)
Critical region+1 (*in*)						
Subject-extraction	510	(128)	425	(108)	410	(73)
Object-extraction	509	(119)	422	(112)	403	(69)

All models were estimated using restricted maximum likelihood. For both the critical region and the subsequent word, the results presented in Table 4 represent the best-fit model using the likelihood ratio statistic (χ^2_{deviance}) to compare the -2 restricted log likelihood values of preliminary versions of the model (Hoffman and Rovine 2007; Quené and van den Bergh 2008).

In the mixed-effect model, Extraction Type (subject-extraction vs. object-extraction) was entered as a fixed effect. The three-way factor of Group (intermediate L2 speakers vs. advanced L2 speakers vs. English native speakers) was recoded into two dummy variables and entered as fixed effects. Both participants and items were entered as random effects.

As seen in Table 4, at both the critical region and the word immediately following the critical region there was a significant effect for Intermediate L2 speakers (Group2) but no significant effect for Advanced L2 speakers (Group1), indicating that the English native speakers were significantly faster overall than the intermediate L2 speakers but that there was no significant difference in overall reading speed between the English native speakers and the advanced L2 speakers. At the critical region there was also an effect for Extraction Type, with longer estimated reading times on subject-extractions relative to object-

Table 4. *Estimated coefficients from the multilevel model analyses of reading time data from all three participant groups (standard errors in parentheses)*

	Critical region			Critical region + 1		
Fixed effects	Coefficient		t	Coefficient		t
(Intercept)	509	(43.7)	11.64****	526	(34.7)	15.16****
Group1	18	(34.8)	0.52	−18	(27.3)	−0.67
Group2	−82	(37.2)	−2.22*	−105	(29.3)	−3.58**
Ext. type	39	(17.2)	2.29*	−7	(19.8)	−0.36
Group1 × Ext. Type	−8	(13.2)	−0.63	4	(15.2)	0.26
Group2 × Ext. Type	−38	(14.8)	−2.56*	10	(17.1)	0.56
Random effects		Variance	*SD*		Variance	*SD*
Participants		15603	125		8749	94
Items		272	16		516	23
Residual		9006	95		11714	108

Estimates are the restricted maximum likelihood. "Group1" and "Group2" refer to participant group (intermediate L2 speakers vs. advanced L2 speakers vs. English native speakers). "Ext. type" refers to extraction type (subject-extraction vs. object-extraction).
* $p < .05$ ** $p < .01$ **** $p < .0001$

extractions. Critically, there was a significant interaction between Intermediate L2 speakers (Group2) and Extraction Type at the critical region. To illustrate the significant Intermediate L2 speaker × Extraction Type interaction, Figure 1 displays the estimated mean reading times at the critical region for subject-extractions and object-extractions as a function of participant group. As illustrated in Figure 1, the intermediate L2 speakers exhibit a large difference in reading times between subject- and object-extractions, whereas the advanced L2 speakers and the English native speakers show no differences in reading times between subject- and object-extractions.

3.2.1. *Secondary analysis.* The initial analysis of the reading time data revealed that the intermediate high school L2 speakers exhibited longer reading times on subject-extractions than object-extractions at the critical region. However, grouping the L2 participants according to education level introduced a certain degree of arbitrariness into this analysis, treating L2 proficiency as a categorical variable rather than more accurately representing this factor as a continuous variable. This point is underscored by participants' self-ratings of L2 proficiency and their scores on the Llex English proficiency task: Even though the mean self-ratings and Llex proficiency scores differed between the two L2 participant groups, there were high-school level participants who rated

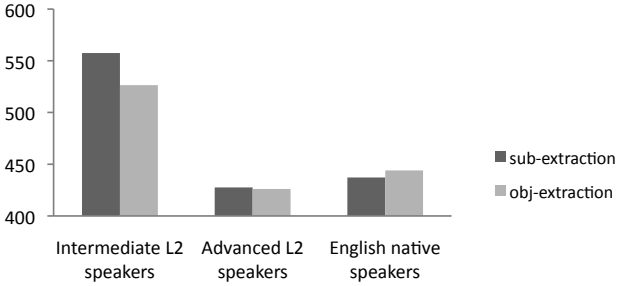

Figure 1. *Estimated reading times in milliseconds for critical region (words 5–7) for all three participant groups*

themselves and achieved Llex scores within the range of the university-level participants and vice versa (see Table 1). To address this shortcoming, a second set of HLM analyses were conducted on the reading time data from just the two L2 participant groups. As in the initial analyses, L2 Participant Group (high school vs. university) and Extraction Type (subject-extraction vs. object-extraction) were entered as fixed effects. L2 English proficiency was also entered as a fixed effect. L2 Proficiency was defined by participants' raw score on the Llex test (with 0 and 100 being the minimum and maximum scores, respectively), which was then centered at the sample mean of 77.59 (Hoffman and Rovine 2007; Kreft et al. 2007). Admittedly, the Llex test only measured participants' lexical knowledge of English and not their understanding of English grammar per se. However, there was a significant positive correlation between participants' Llex scores and their judgment accuracy rates on both the experimental items ($R^2 = .393$, $p = .002$) and the filler items ($R^2 = .682$, $p < .001$).[7] As in the initial analyses, participants and items were entered as random effects.

The 3-way interaction between L2 Participant Group, L2 Proficiency and Extraction Type did not significantly improve the fit of the model and, therefore, was removed from both models (all ps $> .4$). In the final models for each

7. There was also a significant positive correlation between participants' age and L2 proficiency, as measured by the Llex test ($R^2 = .599$, $p < .001$). Therefore, an additional HLM analysis was run including Age, L2 Proficiency and Extraction Type as fixed effects. This analysis revealed no main effect for Age or significant interaction between Age and any other variable (all ps $> .4$), while the interaction between L2 Proficiency and Extraction Type remained significant ($p = .038$). While this does not rule out the possibility that age was an additional factor influencing how the L2 participants processed the target sentences, it suggests that age was not the only factor influencing the L2 participants' performance on the self-paced reading task.

Table 5. *Estimated coefficients from the multilevel model analyses of reading time data (standard errors in parentheses): Dutch–English L2 speakers and L2 proficiency*

	Critical region			Critical region + 1		
Fixed effects	Coefficient		t	Coefficient		t
(Intercept)	447	(37.4)	12.0****	439	(30.1)	14.59****
Group	100	(51.6)	1.93[a]	75	(41.5)	1.80[a]
L2 proficiency	−3	(3.5)	−0.76	−2	(2.8)	−0.83
Ext. type	15	(10.6)	1.43	2	(13.2)	0.14
Group × L2 Prof.	5	(4.1)	1.15	3	(3.2)	0.91
Group × Ext. Type	−1	(19.2)	−0.06	−4	(23.9)	−0.18
L2 prof. × Ext. Type	−2	(0.7)	−2.41*	0	(0.9)	0.07
Group × L2 Prof. × Ext. Type	—		—	—		—
Random effects	Variance		SD	Variance		SD
Participants	18527		136	10828		104
Items	418		20	553		24
Residual	9690		98	14391		120

Estimates are the restricted maximum likelihood. "Group" refers to participant group (high school vs. university). L2 proficiency refers to participants' Llex scores centered on the group mean. "Ext. type" refers to extraction type (subject-extraction vs. object-extraction)

a $p < .1$
* $p < .05$
**** $p < .0001$

region, reported in Table 5, there was a significant effect for L2 Participant Group at the critical region that approached significance on the word following the critical region because the intermediate high-school participants were slower overall than the advanced university participants. While there was no significant interaction between Group and L2 Proficiency or Group and Extraction Type, there was a significant interaction between L2 Proficiency and Extraction Type.

To illustrate the significant L2 Proficiency x Extraction Type interaction, Figure 2 displays the estimated mean reading times at the critical region for subject-extractions and object-extractions as a function of L2 proficiency. In Figure 2, high and low L2 proficiency represent one standard deviation above and below the centered mean. Thus, low L2 proficiency represents estimated reading times from the less-proficient L2 participants and high L2 proficiency represents the estimated mean reading times from the more-proficient L2 participants. The overall pattern shows that the difference in reading times be-

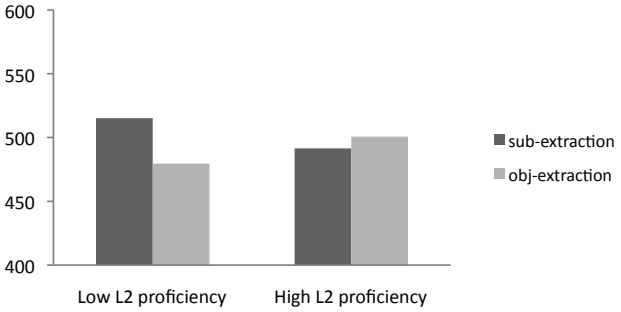

Figure 2. *Estimated reading times in milliseconds for critical region (words 5–7) for Dutch-English L2 speakers according to L2 proficiency*

tween subject- and object-extractions diminishes with increased L2 proficiency, paralleling the reading time patterns for subject- and object-extractions found in the original analyses comparing all three participant groups to each other. Importantly, this interaction between L2 proficiency and extraction type appeared even when L2 participant group was simultaneously entered as a factor in the model, suggesting that the differences in the reading time patterns in the original analyses were driven less by whether participants were in high school or at the university per se, but rather L2 English proficiency.

4. General discussion

To summarize the major findings, the analysis of the reading time data revealed that English native speakers do not necessarily exhibit measureable signs of on-line reanalysis when reading subject- versus object-extractions in English, as there was no significant difference in reading times at the complement clause as a function of extraction type. Similarly, the advanced Dutch–English L2 speakers did not exhibit any significant difference in reading times at the disambiguating complement clause. However, the intermediate Dutch–English L2 speakers did exhibit the predicted processing asymmetry on subject- versus object-extractions in their reading times at the disambiguating complement clause, similar to earlier L2 studies involving the same type of sentences (Dussias and Piñar 2010; Juffs 2005; Juffs and Harrington 1995). Secondary analyses indicated that this processing asymmetry is best explained by L2 proficiency, as opposed to education level per se.

Turning first to the results from the intermediate Dutch–English L2 speakers, their on-line reading times at the disambiguating region show that they had greater difficulty processing subject-extractions than object-extractions when

reading L2 English sentences. These results parallel previously reported processing asymmetries found among L2 speakers of English (Dussias and Piñar 2010; Juffs 2005; Juffs and Harrington 1995), suggesting that the less-proficient Dutch–English L2 speakers suffered some sort of processing breakdown when reading the target sentences. One possible explanation for the intermediate L2 speakers' increased processing costs on subject-extractions is that they had difficulty re-assigning thematic roles to the *wh*-phrase upon discovering that *who* was the subject of the complement clause rather than the direct object of the main clause, whereas no such thematic role re-assignment was necessary on object-extractions (see Dussias and Piñar 2010; Juffs and Harrington 1995). Alternatively, the longer reading times on subject-extractions than object-extractions at the complement clause may have arisen from processing two adjacent finite verbs in the subject-extraction condition (see Juffs 2005). Although the present findings cannot adjudicate between these alternative explanations, the fact that a processing asymmetry between subject- and object-extractions appeared in their on-line results suggests that these intermediate L2 speakers actively searched for a way to integrate the initial *wh*-phrase into the target sentences, similar to findings from previous studies involving both native and highly proficient L2 speakers (e.g., Dussias and Piñar 2010; Jackson and Dussias 2009; Juffs 2005; Juffs and Harrington 1995; Williams 2006; Williams et al. 2001).

At the same time, one must be cautious in interpreting the present findings as evidence that the intermediate L2 speakers necessarily adopted a structurally-based gap-filling strategy when processing the target sentences. For example, their longer reading times at the complement clause on subject-extractions could reflect more generalized difficulties associated with integrating the *wh*-phrase *who* with its subcategorizing verb, rather than difficulties stemming from syntactic reanalysis per se (e.g., Clahsen and Felser 2006; Marinis et al. 2005; Roberts and Felser 2007). Nevertheless, these findings suggest that even less-proficient L2 speakers process L2 input in an incremental manner and, as a result, can exhibit on-line garden path effects when such processing is disrupted.

In contrast to the results from the intermediate Dutch-English L2 speakers, no such processing asymmetries appeared in the analysis of the reading time data from the advanced Dutch-English L2 speakers or the English native speakers. There was also no statistical difference in grammaticality judgment accuracy on either the target *wh*-extractions or the filler items between the advanced Dutch-English L2 speakers and the English native speakers. Further, at least at the descriptive level, the advanced Dutch-English L2 speakers' reading times at both the critical region and on segments prior to the critical region were numerically faster than the English native speakers. One cannot assume that because the advanced Dutch-English L2 speakers did not differ from the

English native speakers in their reading times according to extraction type or in their overall reading speed, that they necessarily processed the target sentences in the same manner as the English native speakers. Nevertheless, the present results do suggest that whatever the source of the processing difficulties among the intermediate Dutch-English L2 speakers, it did not impose a large enough burden upon the advanced Dutch-English L2 speakers or the English native speakers to result in a parallel increase in reading times or decrease in grammaticality judgment accuracy on subject-extraction sentences, similar to results reported by Juffs and Harrington (1995) for English native speakers.

Taken together, the results indicate that, akin to previous studies (Dussias and Piñar 2010; Juffs 2005; Juffs and Harrington 1995), there are real processing costs associated with this type of *wh*-question in English, as evidenced by the on-line results from the intermediate Dutch-English L2 speakers. However, as demonstrated by the results from the advanced L2 speakers, L2 speakers can overcome such difficulties, leading to on-line reading times that do not reveal measurable processing costs for this type of construction. At the same time, due to word order differences in the disambiguating complement clause (e.g., ... *met the tourists* vs. ... *the tourists met*), it is impossible to capture participants' word-by-word processing of subject- versus object-extractions in the present study, as such an analysis would require making direct comparisons across word classes. Therefore, future work comparing on-line processing among intermediate and advanced L2 speakers should investigate other types of *wh*-dependencies or filler-gap constructions, perhaps relying on a more sensitive measure of on-line processing, such as eye-tracking or event-related potentials (ERP) (e.g., Osterhout et al. 2006; Tokowicz and MacWhinney 2005). Doing so could help to determine the precise locus of the observed processing difficulties among intermediate L2 speakers in the present study, as well as uncover more subtle ways in which filler-gap constructions may impose measurable processing costs among more advanced L2 speakers and native speakers.

4.1. *Conclusion*

By examining how both advanced and intermediate Dutch-English L2 speakers parse *wh*-questions in English, the present study underscores the importance of considering proficiency level as a factor influencing on-line processing among L2 speakers. The reading times among the intermediate Dutch-English L2 speakers parallel previous studies that have found on-line processing difficulties for subject-extractions among L2 speakers (Dussias and Piñar 2010; Juffs 2005; Juffs and Harrington 1995). Similar to conclusions drawn by Dekydtspotter et al. (2008), this suggests that even less-proficient L2 speakers will attempt to piece together and integrate sentential arguments in an incremental

manner when reading L2 sentences, leading to increased reading times when such processing breaks down because later input does not match initial expectations. At the same time, the lack of significant reanalysis effects among the advanced Dutch-English L2 speakers and the English native speakers highlights that this type of on-line reanalysis need not always be costly (see also Pickering and Traxler 2001; Staub 2007). As such, the reading time results from the advanced Dutch-English L2 speakers, combined with their high level of judgment accuracy on both the target sentences and on the filler items, demonstrate that at least under certain circumstances, L2 speakers – both off-line and on-line – can functionally behave like native speakers of a language. When taken together, the present findings highlight the importance of investigating how both less- and more-proficient L2 speakers process particular constructions in their non-native language, as findings from each end of the L2 proficiency spectrum can complement each other and help provide a more detailed picture of how L2 speakers process and interpret L2 input in real time.

Penn State University
⟨cnj1@psu.edu⟩
Penn State University/Radboud University Nijmegen
⟨jgv3@psu.edu⟩

Appendix: Example filler items

Grammatical wh-questions (grammatical in English; grammatical in Dutch)

1. *Who does Emily want to invite to her party?*

Ungrammatical wh-questions (ungrammatical in English; ungrammatical in Dutch)

2. **What does he think the lawyer eat for lunch?*

Dependent clause; subject-verb order (grammatical in English; ungrammatical in Dutch)

3. *I know that the boy has read the assignment.*

Dependent clause; verb-final (ungrammatical in English; grammatical in Dutch)

4. **It is tiring that we every day to school must go.*

Split prepositions (grammatical in English; ungrammatical in Dutch)

5. *Who did the child point to in the museum?*

Particle verb; word order (ungrammatical in English; grammatical in Dutch)

6. **The teenagers go every Saturday out.*

Particle verb; incorrect particle (ungrammatical in English; ungrammatical in Dutch)

7. **The thief ate back the banana because he was hungry.*

References

Aarts, Flor. 1994. Relative *who* and *whom*: Prescriptive rules and linguistic reality. *American Speech* 69(1). 71–79.

Baayen, R. H., D. J. Davidson, and D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412.

Bader, Markus & Michael Meng. 1999. Subject-object ambiguities in German embedded clauses: An across-the-board comparison. *Journal of Psycholinguistic Research* 28(2). 121–143.

Carlson, Greg N. & Michael K. Tanenhaus. 1988. Thematic roles and language comprehension. In Wendy Wilkins (ed.), *Syntax and semantics: Thematic relations,* 263–288. San Diego, CA: Academic Press.

Clahsen, Harald & Claudia Felser. 2006. Grammatical processing in language learners. *Applied Psycholinguistics* 27(1). 3–42.

Cuetos, Fernando & Don C. Mitchell. 1988. Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition* 30(1). 73–105.

Dekydtspotter, Laurent, Bryan Donaldson, Amanda C. Edmonds, Audrey L. Fultz & Rebecca A Petrusch. 2008. Syntactic and prosodic computations in the resolution of relative clause attachment ambiguity by English-French learners. *Studies in Second Language Acquisition* 30(4). 453–480.

Dekydtspotter, Laurent, Bonnie Schwartz & Rex A. Sprouse. (2006). The comparative fallacy in L2 processing research. In Mary Grantham O'Brien, Christine Shea & John Archibald (eds.), *Proceedings from the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)*, 33–40. Somerville, MA: Cascadilla Proceedings Project.

Dijkstra, Ton. 2005. Bilingual visual word recognition and lexical access. In Judith F. Kroll & Annette M.B. de Groot (eds.), *Handbook of bilingualism*, 179–201. Oxford, England: Oxford University Press.

De Vincenzi, Marica. 2000. Cross-linguistic psycholinguistics. In Matthew W. Crocker, Martin Pickering & Charles Clifton, Jr. (eds.) *Architectures and mechanisms for language processing,* 282–300. Cambridge: Cambridge University Press.

Dussias, Paola E. & Pilar Piñar. 2010. Effects of reading span and plausibility in the reanalysis of *wh*-gaps by Chinese-English L2 speakers. *Second Language Research* 26(4). 443–472.

Felser, Claudia, Leah Roberts, Theodore Marinis & Rebecca Gross. 2003. The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics* 24(3). 453–489.

Fodor, Janet D. 1993. Empty categories in sentence processing: A question of visibility. In Gerry Altmann & Richard Shillcock (eds.), *Cognitive models of speech processing*, 351–400. Hillsdale: Lawrence Erlbaum Associates.

Frazier, Lyn. 1987. Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory* 5(4). 519–559.

Frazier, Lyn & Charles Clifton Jr. 1996. *Construal*. Cambridge, MA: MIT Press.

Frazier, Lyn & Giovanni B. Flores d'Arcais. 1989. Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language* 28(3). 331–344.

Frenck-Mestre, Cheryl. 1997. Examining second language reading: An on-line look. In Antonella Sorace, Caroline Heycock & Richard Schillcock (eds.), *Proceedings of the GALA 1997 Conference on Language Acquisition,* 474–478. Edinburgh: Human Communications Research Center.

Frenck-Mestre, Cheryl. 2002. An on-line look at sentence processing in the second language. In Roberto R. Heredia & Jeanette Altaribba (eds.), *Bilingual sentence processing*, 217–236. Amsterdam: Elsevier.

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68(1). 1–76.

Harley, Birgit & Doug Hart. 1997. Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition* 19(3). 379–400.

Havik, Else, Leah Roberts, Roeland van Hout, Robert Schreuder & Marco Haverkort. 2009. Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning* 59(1). 73–112.

Hermans, Dan, Theo Bongaerts, Kees de Bot & Robert Schreuder. 1998. Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition* 1(3). 213–229.

Hoffman, Lesa & Michael J. Rovine. 2007. Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavioral Research Methods* 39(1). 101–117.

Hopp, Holger. 2006. Syntactic features and reanalysis in near-native processing. *Second Language Research* 22(3). 369–397.

Jackson, Carrie N. 2008. Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning* 54(4). 875–909.

Jackson, Carrie N. & Paola E. Dussias. 2009. Cross-linguistic differences and their impact on L2 sentence processing. *Bilingualism: Language and Cognition* 12(1). 65–82.

Juffs, Alan. 2005. The influence of first language on the processing of *wh*-movement in English as a second language. *Second Language Research* 21(2). 121–151.

Juffs, Alan & Michael Harrington. 1995. Parsing effects in second language sentence processing: Subject and object asymmetries in *wh*-extraction. *Studies in Second Language Acquisition* 17(4). 482–516.

Just, Marcel A., Patricia A. Carpenter & Jacqueline D. Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111(2). 228–238.

Kaan, Edith. 1997. *Processing subject-object ambiguities in Dutch*. Groningen, NL: Rijksuniversiteit Groningen dissertation.

Keating, Gregory D. 2009. Sensitivity to violations in gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning* 59(3). 503–535.

Kreft, Ita G.G., Jan de Leeuw & Leona S. Aiken. 2007. The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research* 30(1). 1–21.

Kroll, Judith F. & Natasha Tokowicz. 2005. Models of bilingual representation and processing: Looking back and to the future. In Judith F. Kroll & Annette M.B. de Groot (eds.), *Handbook of bilingualism,* 531–553. Oxford, England: Oxford University Press.

Lee, Jin-Hwa. 2010. A subject-object asymmetry in the comprehension of *wh*-questions by Korean learners of English. *Applied Linguistics* 31(1). 136–155.

Mady, Callie. 2007. The suitability of core French for recently arrived adolescent immigrants to Canada. *Canadian Journal of Applied Linguistics* 10(2). 177–196.

Marinis, Theodore, Leah Roberts, Claudia Felser & Harald Clahsen. 2005. Gaps in second language processing. *Studies in Second Language Acquisition* 27(1). 53–78.

Meara, Paul. 1994. Llex: Lingua Vocabulary Tests. Swansea: Centre for Applied Language Studies, University of Wales.

Meara, Paul & Barbara Buxton. 1987. An alternative to multiple choice vocabulary tests. *Language Testing* 4(2). 142–151.

Osterhout, Lee, Judith McLaughlin, Ilona Pitkänen, Cheryl Frenck-Mestre & Nicola Molinaro. 2006. Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning* 56(S1). 199–230.

Pickering, Martin J. & Matthew J. Traxler. 2001. Strategies for processing unbounded dependencies: Lexical information and verb-argument assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27(6). 1401–1410.

Pritchett, Bradley L. 1992. *Grammatical competence and parsing performance*. Chicago: The University of Chicago Press.

Quené, Hugo & Huug van den Bergh. 2008. Examples of mixed-effects modeling with crossed random effects and with binominal data. *Journal of Memory and Language* 59(4). 413–425.

Roberts, Leah & Claudia Felser. 2007. Processing *wh*-dependencies in a second language: A cross-modal priming study. *Second Language Research* 23(1). 9–36.

Schneider, Walter, Amy Eschmann & Anthony Zuccolotto. 2002. E-Prime v1.1. Pittsburgh, PA: Psychology Software Tools Inc.

Staub, Adrian. 2007. The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33(3). 550–569.

Tokowicz, Natasha & Brian MacWhinney. 2005. Implicit and explicit measures of sensitivity to violations in second language grammar. *Studies in Second Language Acquisition* 27(2). 173–204

Traxler, Matthew J., Robin K. Morris & Rachel E. Seely. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language* 47(1). 69–90.

van Hell, Janet G. & Natasha Tokowicz. (2010). Event-related potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research* 26(1). 43–74.

von Studnitz, Roswitha E. & David W. Green. 2002. Interlingual homograph interference in German–English bilinguals: Its modulation and locus of control. *Bilingualism: Language and Cognition* 5(1). 1–23.

Williams, John. 2006. Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition* 9(1). 71–81.

Williams, John N., Peter Möbius & Choonkyong Kim. 2001. Native and non-native processing of English *wh*-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics* 22(4). 509–540.

Zagar, Daniel, Joel Pynte & Sylvie Rativeau. 1997. Evidence for early closure attachment on first-pass reading times in French. *Quarterly Journal of Experimental Psychology* 50A(2). 421–438.