The processing of garden-path sentences by L2 learners of English: a visual word study

Carla Contemori[a], Lucia Pozzan[b], Phillip Galinsky[c] & Paola E. Dussias[c]

[a]University of Texas, El Paso; [b]University of Pennsylvania; [c]Pennsylvania State University

Abstract

We report the results of a study investigating how bilinguals resolve PP-attachment ambiguity and whether they take into account relevant contextual information to aid in the process of ambiguity resolution. The eye movements of highly proficient Spanish-English bilinguals were recorded while they listened to sentences containing PP ambiguities ('put the frog on the napkin into the box') and unambiguous controls ('put the frog that's on the napkin into the box') and performed the actions described in the sentences by moving the objects displayed on a computer screen. Eye-movement records were compared to a group of monolingual English speakers.

The findings showed an increased processing cost associated with the revision of ambiguous sentences for bilinguals relative to monolinguals. In addition, when the discourse was informative, bilinguals were able to use it to recover from an initial misinterpretation. When discourse information was not available, bilinguals pursued a pragmatically infelicitous interpretation of the temporally ambiguous syntactic structure more often than the monolinguals.

Our results suggest that monolingual and bilingual processing is similar with regards to sensitivity to discourse information, but different with regards to reanalysis in the absence of salient contextual information.

Keywords: sentence processing, garden-path sentences, bilingualism, referential information.

Introduction

A central question in psycholinguistic research on bilingual sentence processing is how the parser integrates different types of information during sentence comprehension. This line of research has indicated that different cues can help syntactic revision in the L2 (morphosyntactic cues: e.g., Frenck-Mestre & Pynte, 1997; Jackson, 2008; semantic cues: e.g., Pan & Felser, 2011, Pan et al., 2015; punctuation: Hopp, 2013). However, very little is known about bilingual's sensitivity to sentence-external interpretation cues during comprehension (Pan & Felser, 2011; Pozzan & Trueswell, accepted).

A theory of bilingual sentence processing called the Interface Hypothesis (IH) has argued that bilinguals have reduced abilities to integrate different types of information in real time (e.g., Sorace, 2005, 2011). Sorace (2005, 2011) proposed that in situations that require increased processing demands, bilinguals may show increased costs associated with the simultaneous integration of different information due to general processing capacity limitations. In particular, phenomena that are dependent on syntactic and pragmatic/contextual variables may be vulnerable in L2 acquisition and may lead to residual optionality even at the highest levels of proficiency. According to the IH, in the case of syntactic structures that are temporally ambiguous we expect that bilinguals cannot integrate pragmatic/contextual information as efficiently as native monolingual speakers, due to processing limitations. In this case scenario, the bilingual parser should adopt the simplest syntactic structure more often than monolingual native speakers, even when the referential context supports a more complex syntactic structure.

In the present study, we address this theory of bilingual processing by investigating how the referential information available in the visual scene is integrated to interpret temporally ambiguous sentences in the non-dominant language of Spanish-English bilinguals. The focus of the study is PP attachment in syntactic structures that are temporally ambiguous. The bilinguals' first language is Spanish, which has a similar PP attachment as English. Therefore, any difference between the monolingual and bilingual group observed in the present study should be due to processing

differences rather than cross-linguistic differences. Furthermore, we select a group of participants who are highly proficient in their L2, to test the hypothesis that a residual difficulty may be observed at the syntax-pragmatics interface in the advanced stages of language acquisition (Sorace, 2005; 2011).

The second aim of the study is to understand which underlying process may be more affected when parsing a proficient L2. To the extent that highly proficient bilinguals show increased performance with respect to less proficient bilinguals, such differences might stem from a number of sources, and it is still an open question which of these sources may cause vulnerability in the bilinguals in comparison to the monolinguals. It may be that the L2's parser is more likely to garden path than the monolingual parser, exhibiting stronger commitment to the simplest syntactic structure from the early stages of processing syntactic ambiguities. Alternatively, the L2 speakers may be less efficient than the L1 speakers in abandoning the faulty representation, and consequently experience a higher processing cost associated with revision. In the present study we address this open question, by investigating the interplay between syntactic structure and referential context in the visual word. We look at early and later effects of syntactic ambiguity during listening, to disentangle the nature of the processing difficulty observed with the bilinguals in the L2. Furthermore, we compare on-line processing and off-line comprehension to understand how successful bilinguals can be in the revision process.

While previous studies have mainly used self-paced reading and eye-tracking during reading to investigate L2 sentence processing, in the present study we analyze the interpretation of temporally ambiguous sentences by measuring off-line comprehension (actions performed by the speakers on objects) and eye-movements during listening, a modality which is still relatively unexplored in bilingual sentence processing. Our design and materials resemble those of previous garden-path studies using the visual word paradigm (Trueswell et al., 2009; Pozzan & Trueswell, accepted).

To our knowledge, only one study has looked at the processing of garden path sentences during listening (Pozzan & Trueswell, accepted), testing intermediate learners of English whose L1 is Italian in an act out task where the listener's eye-movements were monitored. Pozzan & Truswell tested their participants in the L1 environment and found that when L2 learners listened to temporarily ambiguous instructions, they were able to use referential information on-line to disambiguate the sentences. However, the L2 learners' off-line accuracy measured with the act-out actions revealed increased difficulties in revising initial interpretations compared to the native speakers. In the present study we ask if proficient bilinguals whose L1 is Spanish experience similar difficulties as the bilinguals with intermediate proficiency in the L2 when tested in their non-dominant language. Differently from Pozzan & Trueswell, we will test bilinguals living in an English-speaking country, thus in a situation of immersion, that we know enhances L2 proficiency more than classroom learning alone (e.g., Talamas, Kroll, & Dufour, 1999).

The first aim of the present study is to provide important information on the parsing and interpretation of temporally ambiguous syntactic structures in bilinguals during listening, testing the integration of referential context and analyzing different stages of processing (early commitments, reanalysis, off-line comprehension). Secondly, by testing bilinguals who are very proficient in the L2, we aim at shedding light on the attainment of the L2 processing routines.

Methods

Design and procedure

Participants listened to temporally ambiguous and unambiguous instructions, as illustrated in (1) and (2), and were asked to perform the action spoken in the instruction using a mouse to move the objects presented on a computer screen.

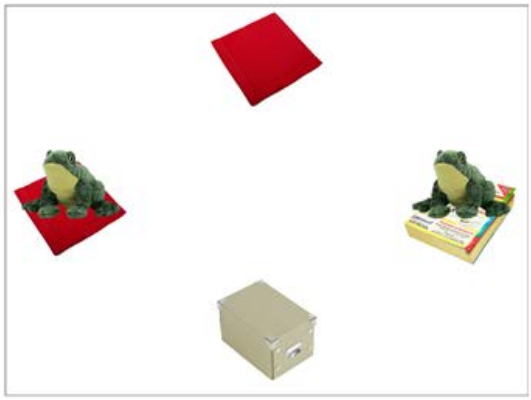
The referential context was manipulated so that when participants heard a temporally ambiguous sentence like (1) they saw either a 2-referent context (see Figure (1A)), or a 1-referent context (see Figure (1B)). We expected to observe more looks to the Incorrect Goal (IG, e.g., the empty napkin) when only one referent (e.g., one frog, as in 1A) is present in the context, because

participants should be more likely to interpret the first PP (on the napkin) as the destination of the action. For the 2-referents context, we do not expect to observe looks to the IG, because the referential context (e.g., the two frogs, as in 2A) provides an early cue to disambiguation, hence in this case the participants should not consider the destination interpretation.

Ambiguity	*Sentence Example*
Ambiguous	(1) Put the frog on the napkin into the box
Unambiguous	(2) Put the frog that's on the napkin into the box

Figure 1. Example of a visual word scene for the Ambiguous 2-Referents context (1A), and Ambiguous 1-Referent context (1B) for the experimental sentence: "Put the frog on the napkin into the box" (source: Pozzan & Truswell, accepted).

(1A)



(1B)

The experimental factors that we manipulated in the task are temporary ambiguity (ambiguous vs. unambiguous) and referential context (1- vs. 2-referents). Each participant heard 2 practice trials, 24 experimental trials (6 per condition), 32 filler continuations for the experimental trials (e.g., "Now move it up and down") and 36 filler trials (e.g., "Put the bowl onto the card") in a pseudo-randomized order. Experimental sentences and fillers were counterbalanced across four experimental lists in a Latin square design.

First, the objects were presented on the screen and given a label (e.g., "a frog", a "napkin", "a box"). Participants looked at the cross at the center of the screen to hear the instruction. After

listening to the instruction, they had 2000 ms to act out the instruction using the mouse and after the 2000 ms they heard a ‘beep”.

The actions were recorded during the experiment using a desktop recorder program. The experimenter transcribed the actions during the experiment and a research assistant unaware of the aim of the experiment checked the transcriptions based on the recordings.

Eye-movements were recorded with a desktop Eyelink 1000 system (SR Research). The eye-tracker monitored the eye position every millisecond, and calibration was based on a 9-point grid.

Participants

Twenty-five English monolingual speakers (mean age: 20 years; SD: 2) and 20 unbalanced bilinguals (mean age: 24 years; SD: 5) participated in the study. One monolingual participant and one bilingual participant were excluded for severe eye-tracking data loss (equal to or higher than 60%).

Monolingual English speakers were undergraduate students at a large US university and received course credits for their participation. The bilinguals were undergraduate and graduate students at the same institution and were compensated for their participation. The bilinguals spoke Spanish as their L1 and English as their L2, and they had been immersed for at least a year in the L2 environment when they were tested. They were all born in a Spanish-speaking country (Central/South America) and moved to the US at different times in their lives. They were first exposed to English at different times during childhood, with some participants having early exposure (Table 1).

The bilinguals were selected on the basis of their proficiency as measured with a subsection of the *Michigan English Language Institute College English Test* (MELICET). The subsection of the MELICET contained 50 multiple-choice questions in two sections, 30 grammar questions and 20 cloze questions from a reading passage. Only bilinguals who scored more than 75% of correct answers on the MELICET participated in the eye-tracking study. In Table 1, we present information

collected with a Language History Questionnaire (Marian, Blumenfeld, & Kaushanskaya, 2007) on the language background of the bilinguals, and their proficiency in the L2 measured with the MELICET.

Table 1. Participant information: Mean (SD)

	Spanish - L1	English - L2
Age of exposure	0 (0)	6 (4)
Length of residence in a country where the language is spoken	16 (7)	7 (6)
Speaking (% average daily)	44 (27)	56 (29)
Reading (% average daily)	35 (24)	65 (24)
Average daily exposure (%)	35 (25)	65 (25)
English Language proficiency (MELICET) Score (out of 50)	-	42.7 (2.9)

Results

Off-line actions

In Figure 2 we show the percentage of actions toward the IG in the ambiguous and unambiguous condition, in the contexts with one and two referents.

We counted actions to the IG that included a movement of a target object (e.g., a frog) to the IG (e.g., the empty napkin). Additionally, we included actions in which the participant dragged the IG onto the Goal object (e.g., the box) and then moved the target object (e.g., a frog) on top of the two.

Figure 2. Proportion of Incorrect Goal actions in the four conditions.

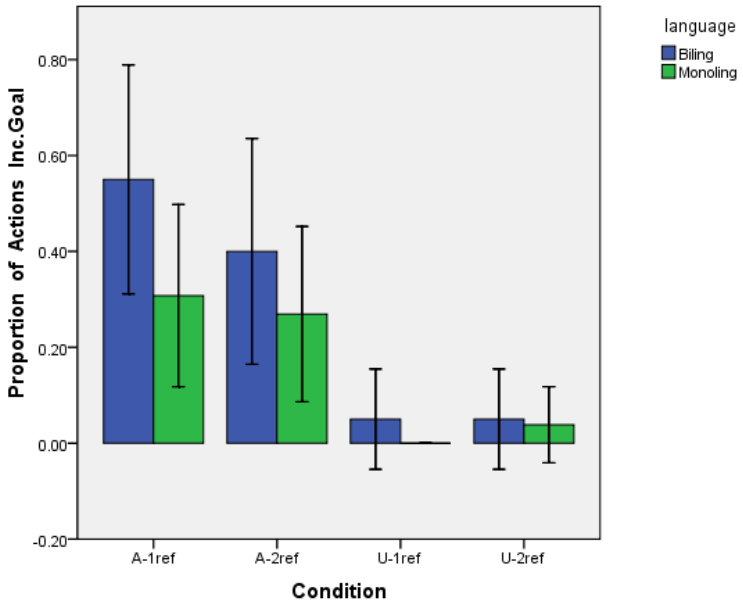


Table 2. Full model statistics for the actions' analysis

Fixed effects:				
	Estimate	Std. Error	t value	pvalue
(Intercept)	.052	0.01	3.9	
Group	.023	0.01	2.36	0.05*
Ref.Context	-.008	0.006	-1.44	0.2
Ambiguity	-.04	0.006	-7.92	0.001***
Group*Ref.Context	-.01	0.005	-2.36	0.01**
Group*Ambiguity	-.02	0.005	-3.41	0.001***
Ref.Context*Ambiguity	.01	0.006	1.87	0.3
Group*Ambiguity*Ref.Context	.01	0.005	2.07	0.05*

Notes
Significance levels throughout the paper: †p<.10; *p<0.05; **p<0.01, ***p<001
The maximal random effect structure leading to convergence includes by subject and by item random intercepts, and by subject random slope.

We used a multi-level linear model to analyze the amount of actions toward the Incorrect Goal using glmer (*lme4* library, Bates & Sakar, 2007). The fixed effects in the model included Ambiguity (ambiguous vs. unambiguous), Referential Context (1 vs. 2 referents) and Group (L1 vs. L2) as main effects, together with all interactions. The results of the full model are presented in Table 2.

The analysis revealed a main effect of Ambiguity, showing more actions to the IG in the ambiguous conditions, and a main effect of Group, indicating more actions to the IG in the bilingual

than in the monolingual group. We also found significant interactions between Group and Ambiguity, Group and Referential Context, and a three-way interaction between Group, Ambiguity and Referential Context.

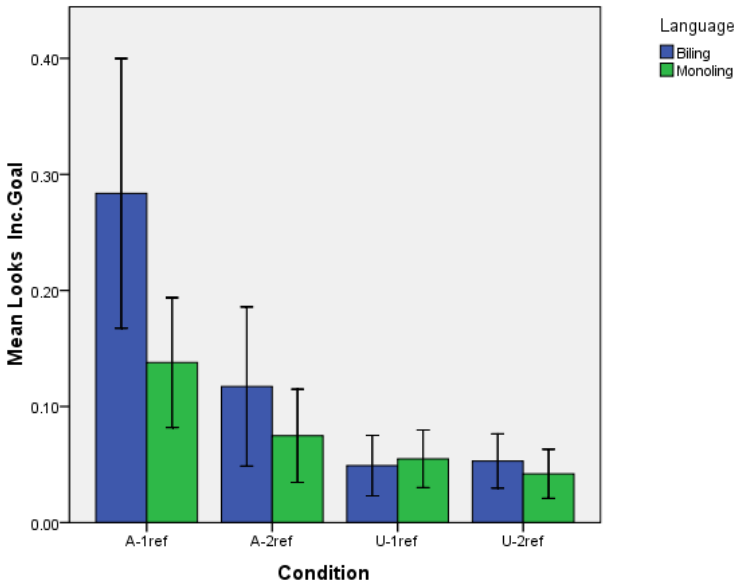
We followed up on the three-way interaction, by comparing the two groups on each condition separately. The planned comparisons showed a significant difference between monolingual and bilinguals on the ambiguous condition with one referent (*Estimate* = 0.06, *SD* = 0.02, t=3.313, p<0.0018), with bilinguals performing significantly more IG actions than the monolinguals. No significant effect emerged for the other conditions (Ambiguous, 2-referents: *Estimate* = 0.01, *SD* = 0.01, t=0.945, p<0.34; Unambiguous, 1-referent: *Estimate* = 0.004, *SD* = 0.003, t=1.149, p<0.2; Unambiguous, 2-referents: *Estimate* = 0.0009, *SD* = 0.005 t=0.182, p<0.8).

Eye-tracking results

In Figure 3 we present the proportion of looks toward the IG in the 5000 ms following the onset of the first PP (e.g., on the napkin).

Trials with combined looking times to the six pictures of less than 30% for the whole time 5000 ms time-window were discarded. These accounted for 5.9% of the trials. Once the trials were removed, the average track loss was 30%.

Figure 3. Mean proportion of looks to IG in the four conditions. Error bars represent 95% Confidence Intervals.



We used a multi-level linear model to analyze the elogit transformed proportion of looks to the Incorrect Goal using lmer (*lme4* library, Bates & Sakar, 2007). Ambiguity (ambiguous vs. unambiguous), Referential Context (1 vs. 2 referents) and Group (L1 vs. L2) were included as main effects. The results of the full model are presented in Table 3.

The analysis showed a main effect of Referent, indicating more looks to the IG in the 1-referent context compared to the 2-referent context, and a main effect of Condition, revealing significantly more looks to the IG in the ambiguous condition compared to the unambiguous condition. We also found significant interactions between Language Group and Condition, and between Referent and Condition. We followed up the Language Group by Condition interaction by comparing the two language groups on the ambiguous and unambiguous conditions separately. The planned comparisons revealed a significant difference between the two groups on ambiguous sentences (*Estimate* = .01, *SD* = .06, t=2.3, $p < .02$), with bilinguals looking significantly longer at the Incorrect Goal than the native speakers. No significant difference emerged between the two groups on the unambiguous conditions (*Estimate* = 1.06, *SD* = 9.05, t=0.012, $p < .99$).

By analyzing the two ambiguous conditions separately, a significant difference between the two groups emerged on the ambiguous condition 1-referent (*Estimate* = .0005, *SD* = .0003, t=2.5, $p < .01$), but not on the ambiguous condition 2- referents (*Estimate* = .0001, *SD* = .0001, t=1.1, $p < .2$), showing that the Ambiguity by Language Group interaction was driven by the ambiguous condition with one referent.

Finally, the absence of a Referent by Language Group interaction suggests that monolinguals and bilinguals' looks to the IG in the two referential contexts are comparable, with both groups showing similar sensitivity to the referential context.

Table 3. Full model statistics for the analysis of looks to the IG (5000 ms following the first PP)

Fixed effects:				
	Estimate	Std. Error	t value	p-value
(Intercept)	.002	.001	-0.85	
Group	.001	.0009	1.93	0.1†
Ref.Context	-.002	.0006	-3.45	0.001***
Ambiguity	-.003	.0006	-4.98	0.01**
Group*Ref.Context	-.006	.0006	-0.87	0.2
Group*Ambiguity	.001	.0006	2.67	0.01**
Ref.Context*Ambiguity	.001	.0006	3.22	0.01*
Group*Ambiguity*Ref. Context	.001	.0006	1.57	0.1

Notes
The maximal random effect structure leading to convergence includes by subject and by item random intercepts.

Even though we found a difference between bilinguals and monolinguals in the online processing of ambiguous sentences, the analysis of the data does not clarify if bilinguals' increased looks to the IG demonstrate an early commitment to the faulty representation (i.e., the first PP is goal) compared to the monolinguals, or if they present a general difficulty in revising the first interpretation. To tease apart these two possible interpretations, we analyzed separately an early time-window time-locked to the onset of the first PP (on the napkin), measuring 2000 ms. Considering that it takes on average 200 ms to plan an eye-movement, this time-window included the first PP (measuring on average 1000 ms) and part of the second PP (onto the box). We used the same multi-level linear model that we used for the 5000 ms time-window. The full model is presented in Table 4.

In this time-window we found a main effect of Referent, indicating overall more looks to the 1-Referent condition compared to the 2-Referents condition, and a main effect of Condition, showing more looks to the IC in the ambiguous conditions compared to the unambiguous conditions. We also found an interaction between Language Group and Referent. However, the planned comparisons on the interaction did not show any difference between the two groups in either the 1-referent context (*Estimate* = 0.0003, *SD* = 0.00, t=0.6, $p < .5$) or the 2-referents context (*Estimate* = 0.0003, *SD* = 0.00, t=0.74, $p < .4$).

The absence of a main effect of Group shows comparable looks to the IG in both monolinguals and bilinguals. Furthermore, the absence of a Group by Referential Context interaction suggests that the two groups make similar use of the referential context in the early stages of ambiguity processing.

Table 4. Full model statistics for the analysis of looks to the IG (2000 ms following the first PP)

Fixed effects:				
	Estimate	Std. Error	t value	p-value
(Intercept)	-0.72	.02	-33.56	0.001***
Group	-0.002	.01	-0.15	0.8
Ref.Context	-0.007	.002	-3.74	0.01**
Ambiguity	-0.009	.002	-4.4	0.001***
Group*Ref.Context	-0.006	.002	-2.35	0.05*
Group*Ambiguity	0.002	.002	1.17	0.2
Ref.Context*Ambiguity	0.008	.002	3.88	0.001***
Group*Ambiguity*Ref.Context	0.002	.002	1.38	0.1

Notes
The maximal random effect structure leading to convergence includes by subject and by item random intercepts and by subject and by item random slopes for the effects of referential context and ambiguity

Discussion

The goal of the present study was to shed light on how proficient bilinguals integrate different types of information when processing temporally ambiguous sentences in the L2 in the auditory modality. By collecting on-line and off-line data, we aimed at understanding if bilinguals experience increased processing costs compared to monolingual speakers, and which stage during the ambiguity resolution process is more affected.

The eye-tracking data revealed that in an early time window (2000 ms after the onset of the first PP) bilinguals and monolinguals' looks to the IG in the ambiguous conditions are comparable, indicating that bilinguals do not engage in the IG interpretation more than the monolinguals when only the first PP has been presented. However, in a larger time-window (5000 ms after the onset of the first PP), we found that bilinguals looked significantly more at the IG compared to the monolingual speakers. We interpret this result as showing increased difficulty in revising the IG

interpretation of a syntactic temporally unambiguous structure relative to monolingual speakers.

The analysis of the actions demonstrated that when only one referent was present in the scene, bilinguals performed significantly more actions towards the IG compared to the monolingual speakers. This finding suggests that in the absence of disambiguating contextual information, bilingual speakers persist with the IG interpretation more often than the monolingual speakers. Conversely, when the context provides an alternative referent, bilinguals are as successful as monolinguals at revising the ambiguous instructions, showing that they can integrate referential information and use it to disambiguate the syntactic structure. These observations confirm that bilinguals' ambiguity resolution preferences can be modulated by the referential context in the off-line comprehension, as demonstrated by the analysis of the actions (see Pozzan & Trueswell, accepted; Pan & Felser, 2011; Pan et a., 2015; for similar results).

For the eye-tracking data, we first showed that the two groups do not differ in the early stages of ambiguity processing (i.e. 2000 ms after the onset of the first PP). However, by looking at a larger time-window, the eye-tracking measures suggest that bilinguals experienced a processing cost associated with the reanalysis of the first (faulty) interpretation. As shown by a separate analysis of the ambiguous conditions in the 5000 ms time-window, the processing cost that emerged was driven by the ambiguous condition with one referent. Therefore, when the context provided an alternative referent, bilinguals were able to integrate the contextual information to perform reanalysis and no processing cost was observed. The on-line results show a similar pattern as the off-line actions, showing that referential context contributes to the processing and the interpretation of ambiguous syntactic structures. These results are not in line with theories of bilingual processing claiming that bilinguals may have general processing limitations parsing the non-dominant language in cases when processing demands are high and different sources of information must be integrated (e.g., Sorace, 2005; 2011). Our data show that when the context provides information to disambiguate the sentence, bilinguals are able to integrate this information to process and successfully interpret syntactic structures that are temporally ambiguous. Conversely, for the

ambiguous sentences in the 1-referent context, difficulties are observed in revising an incorrect goal interpretation during online processing and off-line interpretation, showing that bilinguals pursue an interpretation that is pragmatically odd significantly more often than the monolinguals.

To conclude, in the present study we were able to observe that higher processing costs can be found in the processing of unbalanced bilinguals who are very proficient in the L2. Detailed analysis of the time-course revealed that bilinguals are not led down the garden path more than the monolinguals, but rather they experience more difficulties with the revision process. Finally, while referential information is integrated successfully in bilinguals and monolinguals alike, the absence of disambiguating information is associated with less efficient revision in the former group compared to the latter.

References

Bates, D.M. & Sarkar, D. (2007). lme4: Linear mixed-effects models using S4 classes, R package version 0.99875- 6.

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

Frenck-Mestre, C. & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *The Quarterly Journal of Experimental Psychology*, 50A, 119-148.

Hopp, H. (2013). The development of L2 morphology. *Second Language Research*, 29, 3-6.

Hopp, H. (2009). The syntax-discourse interface in near-native L2 acquisition: Off-line and on-line performance. *Bilingualism: Language and Cognition*, 12, 463-483.

Jackson, C.N. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning*, 58(4), 875-909.

Jacob, G. & Felser, C. (2015). Reanalysis and semantic persistence in native and non-native garden path recovery. *Quarterly Journal of Experimental Psychology*.

Marian, V., Blumenfeld, H.K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50, 940-967.

Pan, H.-Y., S. Schimke & C. Felser (2015). Referential context effects in non-native relative clause ambiguity resolution. *International Journal of Bilingualism,* 19(3), 298–313.

Pan, H.-Y. & C. Felser (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua,* 121, 221–236.

Pozzan, L., & Trueswell, J.C. (accepted). Second language processing and revision of garden-path sentences: a visual word study. *Bilingualism: Language and Cognition*.

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, 32, 299–331.

Sorace, A. (2011). Pinning down the concept of 'interface' in bilingualism. *Linguistic Approaches to Bilingualism*, 1, 1–33.

Sorace, A. (2005). Syntactic optionality at interfaces**.** In L. Cornips and K. Corrigan (eds). *Syntax and Variation: Reconciling the Biological and the Social*, pp. 46-111. Amsterdam: John Benjamins.

Talamas, A., Kroll, J.F., Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition*, 2, 45–58.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition*, 73, 89–134.