Bilingualism: Language and Cognition

cambridge.org/bil

Reanalysis processes in non-native sentence comprehension

Hiroki Fujita and Ian Cunnings

University of Reading, Reading, UK

Research Article

Cite this article: Fujita H, Cunnings I (2021). Reanalysis processes in non-native sentence comprehension. *Bilingualism: Language and Cognition* **24**, 628–641. https://doi.org/10.1017/S1366728921000195

Received: 6 August 2019
Revised: 19 March 2021
Accepted: 26 March 2021
First published online: 21 April 2021

Keywords:
Non-native sentence processing; syntactic ambiguity; eye-movements; reading

Address for correspondence:
Dr Hiroki Fujita,
Email: hiroki.fujita@reading.ac.uk

Abstract

We report two offline and two eye-movement experiments examining non-native (L2) sentence processing during and after reanalysis of temporarily ambiguous sentences like "While Mary dressed the baby laughed happily". Such sentences cause reanalysis at the main clause verb ("laughed"), as the temporarily ambiguous noun phrase ("the baby") may initially be misanalysed as the direct object of the subordinate clause verb ("dressed"). The offline experiments revealed that L2ers have difficulty reanalysing temporarily ambiguous sentences with a greater persistence of the initially assigned misinterpretation than native (L1) speakers. In the eye-movement experiments, we found that L2ers complete reanalysis similarly to L1ers but fail to fully erase the memory trace of the initially assigned interpretation. Our results suggested that the source of L2 reanalysis difficulty is a failure to erase the initially assigned misinterpretation from memory rather than a failure to conduct syntactic reanalysis.

Introduction

Syntactic ambiguity resolution has played an important role in motivating research in native (L1) and non-native (L2) sentence processing. Previous studies have shown that L1 and L2 speakers encounter difficulty when reading temporarily ambiguous sentences (e.g., Frazier & Rayner, 1982; Juffs & Harrington, 1996). For example, in (1), a temporary ambiguity emerges at "the baby", which can be interpreted as either the subordinate clause object or the main clause subject. Although the latter is the globally correct interpretation, the former interpretation ("Mary dressed the baby") may be initially adopted. This initial misinterpretation requires reanalysis later in the sentence, at the disambiguating verb ("laughed").

(1) While Mary dressed the baby laughed happily.

Importantly, research has shown that the initial misinterpretation may persist after reanalysis (Christianson, Hollingworth, Halliwell & Ferreira, 2001), and that L2ers may have increased reanalysis difficulty (Pozzan & Trueswell, 2016; Jacob & Felser, 2016). However, what causes this L2 reanalysis difficulty is debated.

To explore this issue, we report four experiments. Two experiments used offline methods to examine the final interpretation assigned to garden-path sentences, while two used eye-tracking during reading to investigate the time-course of reanalysis. Our results suggest that L1 and L2 reanalysis difficulty resides mostly in a difficulty in discarding the initial misinterpretation from memory, rather than an inability to conduct syntactic reanalysis. Below, we begin by discussing approaches to reanalysis and misinterpretation in L1ers, before discussing potential differences between L1 and L2 processing.

Reanalysis in L1 sentence processing

Many studies have shown that sentences like (1) cause reading difficulty at "laughed" (e.g., Sturt, Pickering & Crocker, 1999). This suggests that readers initially misinterpret "the baby" as the theme of "dressed", and subsequently attempt to correct this misinterpretation. Although different theories account for why the misinterpretation is initially considered in different ways, either due to an initial parsing preference for the syntactically simplest structure (e.g., Frazier & Rayner, 1982) or because multiple interacting constraints support it (e.g., MacDonald, Pearlmutter & Seidenberg, 1994), all assume that in sentences like (1), "the baby" is initially misinterpreted as the theme of "dressed". Additionally, models differ in terms of whether a single or multiple analyses of an ambiguous input are constructed. Serial models posit that a single structure is initially computed that later needs to be revised (e.g., Frazier & Rayner, 1982; Van Gompel, Pickering & Traxler, 2000). In parallel models, multiple different structures are computed in parallel, with garden-path sentences requiring a reranking of the different possibilities once disambiguating information is encountered (e.g., Gibson, 1991). We do not attempt to tease apart these different accounts. For present

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

CAMBRIDGE UNIVERSITY PRESS

purposes, all models assume that garden-path sentences are initially misinterpreted, and subsequently reinterpreted. We refer to this process as reanalysis.

The difficulty associated with garden-path sentences like (1) indicates that readers have noticed a conflict in the input, and presumably made some attempt to conduct reanalysis. However, garden-path effects by themselves do not necessarily index the degree of reanalysis that is undertaken. Recent studies have examined this issue. Christianson et al. (2001) used end-of-sentence questions (e.g., "Did Mary dress the baby?") that probed interpretation of the ambiguous phrase. Although the correct answer to such questions is "no", they observed more incorrect "yes" responses when questions followed temporarily ambiguous sentences like (1) than unambiguous sentences disambiguated with a comma (e.g., While Mary dressed, the baby laughed happily.). They interpreted this as evidence that the initial misinterpretation ("Mary dressed the baby") often persists following reanalysis.

One counterargument to this claim is that lower accuracy rates for ambiguous sentences may be an artefact of the design, due to reactivation of the misinterpretation as a result of the comprehension question being more similar to the ambiguous than unambiguous sentences (Tabor, Galantucci & Richardson, 2004). However, corroborating results have been found with various designs that avoid such repetition (e.g., Malyutina & den Ouden, 2016; van Gompel, Pickering, Pearson & Jacob, 2006).

The Good Enough account of language comprehension predicts lingering misinterpretation based on how readers process language (e.g., Christianson et al., 2001; Ferreira, Christianson & Hollingworth, 2001). However, the precise nature of how processing is "good-enough" has been debated (see Christianson et al., 2001; Kaschak & Glenberg, 2004; Slattery, Sturt, Christianson, Yoshida & Ferreira, 2013). One possibility might be that readers do not conduct syntactic reanalysis. This account predicts that lingering misinterpretation arises due to a failure to construct the correct syntactic structure. That is, in sentences like (1), the temporarily ambiguous noun phrase remains as the subordinate clause object, rather than main clause subject. Alternatively, syntactic reanalysis may be conducted, but lingering effects arise because the initially assigned misinterpretation is not fully discarded. In this way, the temporarily ambiguous noun phrase in (1) is reanalysed as the main clause subject, but the initial misinterpretation lingers in memory.

To tease these two accounts apart, Slattery et al. examined misinterpretation in two eye-movement experiments. In one experiment, participants read sentences like (2). Sentences were either temporarily ambiguous or made unambiguous by including a comma, and additionally manipulated the gender (mis)match between a reflexive ("himself") and its antecedent ("father/mother").

(2) After the bank manager telephoned(,) David's father/mother grew worried and gave himself approximately five days to reply.

Slattery et al. reported longer reading times at "grew" when the comma was absent, but they were particularly interested in subsequent processing at the reflexive. Following Sturt (2003), for unambiguous sentences, longer reading times were predicted when the reflexive mismatched in gender with its antecedent ("David's mother grew worried and gave himself…") compared to when it matched ("David's father… himself"). Slattery et al. reasoned that if syntactic reanalysis is not conducted in ambiguous sentences, such that "David's father/mother" remains as the subordinate clause object rather than main clause subject, gender mismatch effects should be reduced or absent in ambiguous conditions. This is because according to Binding Principle A (Chomsky, 1981), the reflexive's antecedent in (2) must be the main clause subject. If reanalysis is incomplete, however, the reflexive may fail to find an antecedent. Contrary to this incomplete syntactic reanalysis hypothesis, Slattery et al. observed gender mismatch effects in both ambiguous and unambiguous sentences, suggesting L1ers perform syntactic reanalysis of the temporarily ambiguous noun phrase.

In a second experiment, Slattery et al. tested texts like (3).

(3) While Frank dried off(,) the truck/grass that was dark green was peed on by a stray dog. Frank quickly finished drying himself off then yelled out the window at the dog.

The first sentence is either temporarily ambiguous or unambiguous. It also manipulates whether the main clause subject is a plausible or implausible theme of the subordinate clause verb (plausible "dried off the truck" vs. implausible "dried off the grass"). Slattery et al. reasoned that for ambiguous sentences, plausible initially assigned misinterpretations may linger, compared with implausible ones.

The second sentence always referred to the globally correct interpretation of the first sentence ("Frank quickly finished drying himself off"). It is, however, inconsistent with the initial misinterpretation ("Frank dried off the truck"). Thus, if the initial misinterpretation is completely erased, there should be no reading time differences between conditions at the reflexive. However, if misinterpretation lingers in the plausible conditions, reading times at "himself" may become longer in ambiguous than unambiguous conditions, as evidence that the initial misinterpretation ("Frank dried off the truck") lingered. Consistent with this latter prediction, Slattery et al. observed longer reading times for ambiguous than unambiguous sentences at the reflexive in plausible conditions. Taking the results of both experiments together, Slattery et al. argued that L1ers conduct syntactic reanalysis of the temporarily ambiguous noun phrase but do not fully erase the initial misinterpretation from memory (see also Kaschak & Glenberg, 2004).

Reanalysis in L2 sentence processing

Previous research has shown that L2ers are garden-pathed like L1ers (e.g., Juffs & Harrington, 1996; for review see Cunnings, 2017). However, some recent studies report that L2ers have more difficulty reanalysing temporary ambiguities than L1ers (e.g., Gerth, Otto, Nam & Felser, 2017; Jacob & Felser, 2016; Pozzan & Trueswell, 2016). For example, Pozzan and Trueswell tested sentences like (4).

(4a) Put the frog on the napkin onto the box.
(4b) Put the frog that's on the napkin onto the box.

(4a) causes garden paths at "onto the box," as the preceding prepositional phrase ("on the napkin") is initially misanalysed as the destination of "put". (4b) is unambiguous due to the overt complementiser "that". Participants heard sentences like (4) while viewing a display containing the referents mentioned in the sentence, and then acted out the instruction. Eye-movements during listening showed L1ers and L2ers similarly misinterpreted the temporarily ambiguous sentences. However, L2ers performed more incorrect actions than L1ers following ambiguous sentences only, often

https://doi.org/10.1017/S1366728921000195 Published online by Cambridge University Press

moving the frog first to the napkin and then the box. This suggests increased reanalysis difficulty for L2ers.

Jacob and Felser (2016) examined L2 reanalysis using an eye-movement while reading task. In their study, participants read sentences like (5).

(5) While the gentleman was eating(,) the burgers were still being reheated in the microwave.

In (5), the main clause ("the burgers were still being reheated") is semantically inconsistent with the initial misinterpretation ("the gentleman was eating the burgers"). Reading times were longer at "were still" in ambiguous sentences, indicating garden-path effects. Garden-path effects were, however, smaller for L2ers than L1ers in later processing stages (e.g., regression path duration, total viewing times), which Jacob and Felser took as indicating that L2ers are more reluctant to initiate or complete reanalysis than L1ers. L2ers were also less accurate than L1ers at answering post-sentence comprehension questions that probed the initial misinterpretation (e.g., "Was the gentleman eating the burgers?"), which provides some evidence that misinterpretations persist more often in L2ers. Note, however, that L2ers had lower accuracy on both ambiguous and unambiguous trials.

While studies thus suggest L2ers may be more persistent with misinterpretation than L1ers, the mechanisms underlying L2 reanalysis have not been fully examined. One potential account is that L2ers do not conduct syntactic reanalysis. This may be compatible with the Shallow Structure Hypothesis (Clahsen & Felser, 2006, 2017), which claims L2ers have difficulty constructing abstract syntactic structures during sentence processing (e.g., Felser, Roberts, Marinis & Gross, 2003). Although the Shallow Structure Hypothesis was not originally formulated to account for L2 reanalysis processes, L2ers may have difficulty constructing the correct syntactic structure after reanalysis as a result of shallow parsing. Alternatively, L2ers may perform syntactic reanalysis like L1ers but have increased difficulty in erasing the initial interpretation from memory (Cunnings, 2017). This predicts that L2ers conduct syntactic reanalysis like L1ers, but that misinterpretations should be more likely to linger during L2 processing.

The present study

Against this background, we report two studies investigating L2 reanalysis. We aimed to tease apart whether L2 reanalysis difficulty relates to incomplete syntactic reanalysis, or difficulty in erasing initially assigned misinterpretations. In the first study, participants completed an offline task (Experiment 1) that used comprehension questions to investigate the final interpretation assigned to garden-path sentences and an online eye-tracking during reading experiment (Experiment 2) to investigate reanalysis during processing. In the second study, participants completed a sentence-picture matching task (Experiment 3) to further investigate how garden-path sentences are interpreted, and an online eye-tracking experiment that investigated persistence of misinterpretation (Experiment 4). For both studies, although we report the offline tasks first, participants completed the eye-tracking task first in a separate experimental session.

Experiment 1

Experiment 1 examined ambiguous (6a) and unambiguous (6b) sentences. Each experimental sentence was followed by one of two comprehension questions, (7a) or (7b), that tested two aspects of reanalysis. Specifically, reanalysis in (6a) involves revising the subordinate clause verb as intransitive, and the temporarily ambiguous noun phrase as the main clause subject. (7a) tested the former while (7b) examined the latter. We avoided yes/no questions that involve repetition of the temporary ambiguity to minimise potential reactivation of the misinterpretation (Tabor et al., 2004). The two answers always denoted a misinterpretation and the globally correct interpretation.

(6a) *Ambiguous*

After the mother dressed the baby in the living room laughed very happily.

(6b) *Unambiguous*

After the mother dressed, the baby in the living room laughed very happily.

(7a) *Subordinate clause question*

What happened? 1. The mother dressed herself 2. The mother dressed the baby

(7b) *Main clause question*

Who laughed very happily? 1. The mother 2. The baby

We expected lower accuracy for ambiguous than unambiguous sentences (Christianson et al., 2001). If the ambiguous noun phrase is reanalysed as the main clause subject but the initial misinterpretation persists, accuracy rates should be higher for ambiguous main clause questions than ambiguous subordinate clause questions. If L2ers conduct syntactic reanalysis but are more likely to fail to discard initial misinterpretations than L1ers, accuracy should be lower for L2ers only in ambiguous subordinate clause questions. However, if L2 reanalysis difficulty relates to incomplete syntactic reanalysis, accuracy for L2ers should be lower in ambiguous main clause questions as well.

Method

Participants

Forty L1 English speakers (5 males, mean age 19; range 18–23) and 40 L2 English speakers (10 males, mean age 25; range 18–43) of various L1 backgrounds,[1] from the University of Reading community participated in Experiment 1. Participants received either course credit or payment for taking part.

L2 participants started learning English in a school environment after age five (mean age of onset 8.6; SD 3.1; range 5–18). They also completed the Oxford Placement Test (OPT) (Allan, 2004). Their mean scores were 72/100 (SD 10.3; range 52–89), showing they were upper intermediate-advanced English language learners.

[1]First languages were Chinese (6), Spanish (5), French (4), Malay (4), Turkish (4), Cantonese (2), Kazakh (2), Polish (2), Swedish (2), Greek (1), Indonesian (1), Italian (1), Norwegian (1), Russian (1), Serbian (1), Thai (1).

https://doi.org/10.1017/S1366728921000195 Published online by Cambridge University Press

Materials

Experimental materials consisted of 24 sets of sentences as in (6) paired with comprehension questions as in (7). The materials manipulated ambiguity such that a comma was either present (6a) or absent (6b). The subordinate clause always utilised reflexive absolute transitive (RAT) or reciprocal verbs (Ferreira & McClure, 1997). RAT verbs, when used intransitively, must be interpreted reflexively. For example, "the mother dressed" in (6b) can only mean "the mother dressed herself". Reciprocal verbs share a similar property when the subject is plural. These verbal properties are crucial in Experiment 1, as they create a situation where there is always only one absolute correct answer to the comprehension questions.

The experiment also contained 72 filler sentences with a variety of syntactic structures, of which two-thirds were followed by a comprehension question. Experimental sentences were constructed with four counterbalanced presentation lists in a Latin Square design. The full set of experimental items used in Experiments 1–4 can be found online as Online Supplement 1.

Procedure

The experiment was administered as a whole-sentence reading comprehension task using Linger (Rohde, 2010). Each trial began with a cross onscreen. Upon pressing the space bar, an entire sentence was shown. After reading the sentence, participants pressed the space bar again, at which point the sentence was replaced with a question containing two options. Participants answered each question by pressing either the "1" key for the first option or the "2" key for the second option. The experimental and filler sentences were pseudo-randomised so that at least two filler sentences always appeared between each experimental sentence. The order of the answers was also randomised to assign the correct and incorrect answers to the two options equally. The experiment began with some practices.

After the experiment, L2 participants completed the OPT. The main experiment took approximately 25 minutes with an additional 25–30 minutes for the OPT.

Data analysis

We analysed accuracy rates in R (R Core Team, 2018) using generalised linear mixed-effects models. Models included sum-coded fixed effects of ambiguity (ambiguous/unambiguous), question type (subordinate clause question/main clause question) and group (L1/L2).

All models were fit using the maximal random effects structure that converged, including by-subject and by-item random intercepts, and random slopes for each within-item and within-subject fixed effect (Barr, Levy, Scheepers& Tily, 2013).[2] For each fixed effect, p values were estimated using the Laplace Approximation. Data and analysis code for all experiments reported here is available at the first author's Open Science Framework webpage (https://osf.io/bt637/).

Results

Average accuracy rates for filler sentences were 93% for both L1 (range 83–100) and L2 (range 79–100) participants. Accuracy rates and a summary of the statistical analysis are provided in Tables 1 and 2.

[2]If the maximal model failed to converge, we first removed the random correlations. If this model still did not converge, the random effect accounting for the least variance was iteratively removed until convergence was achieved.

There was a significant main effect of ambiguity due to lower accuracy rates for the ambiguous than unambiguous conditions, and a significant main effect of question type, with more correct responses to main clause than subordinate clause questions. There was also a significant three-way interaction between ambiguity, question type and group. Planned 2×2 analyses on each question type showed significantly lower accuracy for ambiguous than unambiguous sentences for subordinate clause questions (*estimate* = 1.792, $z = 5.71$, $SE = 0.31$, $p < .001$). Neither the main effect of group (*estimate* = 0.537, $z = 1.53$, $SE = 0.35$, $p = .127$) or interaction (*estimate* = 0.110, $z = 0.20$, $SE = 0.55$, $p = .840$) was significant. Regarding main clause questions, the main effect of ambiguity but not group was significant (ambiguity: *estimate* = 2.036, $z = 3.36$, $SE = 0.61$, $p < .001$; group: *estimate* = 0.224, $z = 0.59$, $SE = 0.38$, $p = .556$). There was also a marginal two-way interaction (*estimate* = 1.333, $z = 1.71$, $SE = 0.78$, $p = .088$). Planned comparisons by ambiguity revealed lower accuracy rates for L2 than L1 participants in the ambiguous condition (*estimate* = 0.881, $z = 2.11$, $SE = 0.42$, $p = .035$) but not in the unambiguous condition (*estimate* = 0.261, $z = 0.42$, $SE = 0.62$, $p = .673$).

Discussion

Experiment 1 showed that L1 and L2 participants persisted with misinterpretation. Although accuracy rates were generally higher for main clause than subordinate clause questions, the low accuracy rates for ambiguous main clause questions may contrast with Slattery et al. (2013) and Christianson et al. (2001), who claimed that L1ers successfully reanalyse the temporarily ambiguous noun phrase as the main clause subject. This issue is discussed in the General Discussion.

Importantly, there was evidence that reanalysis was more costly for L2ers than L1ers. However, this was observed in main clause but not subordinate clause questions. This is unexpected both from the perspective of L2ers failing to conduct syntactic reanalysis, which would predict lower accuracy in both question types, and from the perspective of L2ers having increased difficulty with lingering misinterpretation, which predicts lower accuracy in main clause questions only. We return to this issue in Experiment 3, but first report Experiment 2 which examines whether L2ers conduct syntactic reanalysis during online reading.

Experiment 2

Experiment 2 investigates reanalysis during sentence processing. We adapted Slattery et al.'s (2013) design. Participants read texts like (8) while their eye-movements were monitored.

(8) Some people had a party at a friend's house at the weekend.

(a) *Ambiguous, Gender Match*
After the neighbour visited Ken's dad decided to prepare himself a cold drink.
(b) *Ambiguous, Gender Mismatch*
After the neighbour visited Ken's mum decided to prepare himself a cold drink.
(c) *Unambiguous, Gender Match*
After the neighbour visited, Ken's dad decided to prepare himself a cold drink.
(d) *Unambiguous, Gender Mismatch*
After the neighbour visited, Ken's mum decided to prepare himself a cold drink.
It was very tasty.

https://doi.org/10.1017/S1366728921000195 Published online by Cambridge University Press

Table 1. Accuracy rates for ambiguous and unambiguous sentences in Experiments 1 and 3 (SDs in parentheses).

	Native Speakers				Non-Native Speakers			
	Subordinate Clause		Main Clause		Subordinate Clause		Main Clause	
	Ambiguous	Unambiguous	Ambiguous	Unambiguous	Ambiguous	Unambiguous	Ambiguous	Unambiguous
Experiment 1								
Accuracy	58 (49)	88 (32)	75 (44)	93 (25)	50 (50)	82 (38)	60 (49)	95 (22)
Experiment 3								
Accuracy	69 (46)	83 (37)	81 (39)	92 (27)	57 (50)	81 (39)	65 (48)	93 (26)

Table 2. Summary of statistical analyses for Experiments 1 and 3.

	Experiment 1			Experiment 3		
	Estimate (SE)	*z value*	*p* value	*Estimate (SE)*	*z value*	*p* value
Ambiguity	2.30 (0.24)	9.71	< .001	1.76 (0.22)	7.94	< .001
Group	0.46 (0.30)	1.55	.120	0.62 (0.33)	1.88	.060
Question Type	0.93 (0.22)	4.24	< .001	0.92 (0.31)	2.95	.003
Ambiguity:Group	0.45 (0.46)	0.97	.331	0.72 (0.39)	1.84	.066
Ambiguity:Question Type	0.31 (0.32)	0.98	.328	0.48 (0.40)	1.21	.228
Group:Question Type	0.25 (0.32)	0.77	.444	0.09 (0.35)	0.26	.797
Ambiguity:Group:Question Type	1.34 (0.64)	2.10	.036	0.56 (0.61)	0.92	.359

(8a/b) are temporarily ambiguous while (8c/d) are unambiguous. We further manipulated gender (mis)match between the reflexive ("himself") and its antecedent ("Ken's dad/mum"). In (8a/c), the antecedent matches the reflexive's gender, whereas in (8b/d), it does not. Garden paths are expected with increased reading times for ambiguous sentences at the disambiguating region ("decided"). If L2ers are more reluctant to initiate or complete reanalysis (Jacob & Felser, 2016), garden-path effects should be smaller for L2ers than L1ers.

In unambiguous sentences, the gender mismatch condition should elicit longer reading times at the reflexive than the gender match condition (Slattery et al., 2013). For the ambiguous conditions, if syntactic reanalysis is conducted, similar gender mismatch effects are expected. However, if syntactic reanalysis is not conducted, such that the temporarily ambiguous noun phrase remains as the subordinate clause verb's theme, gender mismatch effects should be reduced or absent in the ambiguous conditions. For L1ers, we expected to replicate Slattery et al. (2013), and observe gender mismatch effects in both ambiguous and unambiguous conditions. If L2ers fail in syntactic reanalysis, gender mismatch effects should emerge only in the unambiguous conditions. Alternatively, if L2 syntactic reanalysis is successful, they should behave like L1ers.

Method

Participants

The participants from Experiment 1 took part in Experiment 2. Experiment 2 was conducted at least one week before Experiment 1.

Materials

Twenty-four sets of experimental sentences as in (8) were created. Each set began with a lead-in sentence, which always appeared on the first line. The critical target sentence appeared across two lines, with the line-break appearing immediately after the time conjunction ("After" in (8)). Half of the experimental sentences employed the masculine reflexive and half the feminine. A third wrap-up sentence, which took up one line onscreen, was inserted on the third line.

Seventy-two fillers were also constructed with various syntactic structures, which took up either two or three lines on the screen. All experimental and two-third of filler texts were followed by a yes/no comprehension question. No question queried the temporary ambiguity or the reflexive's antecedent.

Procedure

Although viewing was binocular, eye-movements were recorded from the right eye using an SR Research Eyelink 1000 at a sample rate of 1000 Hz. Before each experimental session, calibration of the eye-tracker was conducted on a nine-point grid, and recalibration was conducted when needed. Before each text was shown, participants fixated on a black square above the first word of the text. Upon fixation of the square, the text appeared. After reading the text, participants pressed a button on a game pad to make the text disappear. Either the next trial then began, or a comprehension question was shown, which participants answered by pressing an appropriate button on a gamepad. Experimental and filler texts were presented in a pseudo-randomised Latin-square design.

https://doi.org/10.1017/S1366728921000195 Published online by Cambridge University Press

After the experimental session, L2 participants looked through a vocabulary list containing vocabulary used for the subordinate clause verb and the main clause subject. Participants ticked a box next to each word that they were unsure of. The entire experiment lasted 40–60 minutes.

Data analysis

Three eye-tracking measures are reported for four regions of text.[3] To test for garden-path effects, we analysed the disambiguating region ("decided"), and a first spillover region containing the words up to but not including the reflexive ("to prepare"). To test for gender mismatch effects, we analysed the reflexive region ("himself"), and a second spillover region that contained the rest of the critical sentence ("a cold drink"). Eye-tracking measures included first pass reading times, the summed duration of all fixations within a region until an eye-movement away from the region, regression path duration, the summed duration of fixations from the first fixation entering a region from the left up until but not including the first fixation in a region to the right, and total viewing times, the summed duration of all fixations within a region. Fixations shorter than 80ms that were within one degree of visual arc of another fixation were merged, and any other fixations shorter than 80ms or longer than 800ms were removed. Further, any region that a participant skipped was removed from data analysis, which affected less than 7% of the L1 data and 3% of L2 data. Trials including vocabulary that the L2ers did not know were also removed, which affected less than 0.1% of the L2 data.

Data analysis used linear mixed-effect models (Baayen, Davidson & Bates, 2008). Each model included log-transformed reading times as the dependent variable. Sum-coded fixed effects included ambiguity (ambiguous/unambiguous), gender (gender match/mismatch) and group (L1/L2). We also included region (disambiguating region/first spillover region or reflexive region/second spillover region) as a fixed effect (see Cunnings & Sturt, 2018). As treating region as a fixed effect involves two non-independent datapoints from the same trial, we also included random intercepts for trial, and by-subject, by-item and by-trial random slopes for region. For each fixed effect, p values were estimated using the Satterthwaite approximation implemented by the lmerTest package (Kuznetsova, Brockhoff & Christensen, 2017). Each model was fit with the maximal random effects structure that converged.

Results

Mean accuracy rates across experimental and filler trials were 89% for L1 participants (range 75–100%) and 88% for L2 participants (range 71–97%). A summary of the reading time data is provided in Table 3, and a full summary of the inferential statistics can be found in Online Supplement 2. For brevity, all statistical models showed a significant main effect of group due to longer reading times for L2 than L1 participants. Also, we do not discuss main effects of region, or group by region interactions, as these are of little theoretical interest unless they interact with another fixed effect.

Disambiguating and 1st spillover regions

For both first pass and regression path times at the disambiguating and first spillover regions, there were significant main effects of ambiguity, with longer reading times for ambiguous than unambiguous sentences. In first pass times there was also a significant ambiguity by region interaction, as garden-path effects were observed at the disambiguating region (*estimate* = 0.091, $t = 4.47$, $SE = 0.02$, $p < .001$) but not the first spillover region (*estimate* = 0.008, $t = 0.38$, $SE = 0.02$, $p = .705$).

In first-pass times the effect of gender and the three-way interaction between ambiguity, gender and region were marginal. Follow-up analyses conducted on each region suggested a marginal interaction between ambiguity and gender only in the disambiguating region (*estimate* = 0.083, $t = 1.82$, $SE = 0.05$, $p = .073$), with apparent gender mismatch effects in the ambiguous conditions only (for ambiguous conditions, *estimate* = 0.071, $t = 2.16$, $SE = 0.03$, $p = .034$; for unambiguous conditions, *estimate* = 0.012, $t = 0.41$, $SE = 0.03$, $p = .682$). The main effect of gender and two-way interaction between ambiguity and gender were also significant in regression path times. Follow-up analyses by ambiguity showed apparent gender mismatch effects only in ambiguous sentences (ambiguous: *estimate* = 0.090, $t = 2.62$, $SE = 0.03$, $p = .009$; unambiguous: *estimate* = 0.001, $t = 0.03$, $SE = 0.03$, $p = .976$). We did not expect gender mismatch effects in these measures, and as they appear before the reflexive, we assume these effects are spurious.[4]

In total viewing times, there were significant main effects of ambiguity, with longer reading times in ambiguous sentences, and gender, with longer reading times in gender mismatch conditions. Gender mismatch effects are expected in this measure, as total times can include reading after the reflexive was encountered. There was a marginal three-way interaction between ambiguity, gender and region. Follow-up analyses by region showed main effects of ambiguity (*estimate* = 0.283, $t = 8.08$, $SE = 0.04$, $p < .001$) and gender (*estimate* = 0.125, $t = 2.88$, $SE = 0.04$, $p = .009$) with a marginal interaction between them (*estimate* = 0.109, $t = 1.77$, $SE = 0.06$, $p = .092$) in the disambiguating region. Pairwise comparisons conducted at the two levels of ambiguity indicated gender mismatch effects only in the ambiguous condition (ambiguous: *estimate* = 0.178, $t = 3.03$, $SE = 0.06$, $p = .006$; unambiguous: *estimate* = 0.072, $t = 1.54$, $SE = 0.05$, $p = .138$). Analysis of the first spillover region indicated main effects of ambiguity (*estimate* = 0.201, $t = 5.84$, $SE = 0.03$, $p < .001$) and gender (*estimate* = 0.148, $t = 3.88$, $SE = 0.04$, $p < .001$) only, due to garden-path and gender mismatch effects respectively.

Reflexive and 2nd spillover regions

At the reflexive and second spillover regions, a significant main effect of gender was observed in all eye-movement measures, with longer reading times in gender mismatch than gender match conditions. This effect in total viewing time is illustrated in Figure 1. Additionally, in first pass reading times, the model showed a significant interaction between ambiguity and group, but pairwise comparisons revealed no main effect of ambiguity in either group (L1: *estimate* = 0.041, $t = 1.65$, $SE = 0.02$, $p = .114$; L2: *estimate* = 0.022, $t = 1.09$, $SE = 0.02$, $p = .276$).

[3]We acknowledge that analysing multiple eye-movement measures inflates the probability of false rejections of the null hypothesis (see von der Malsburg & Angele, 2017). We report p values unadjusted for multiple comparisons, but note that if we did adjust alpha by the three measures that we report (.05 / 3 = adjusted alpha .016) our main findings in Experiments 2 and 4 would remain significant.

[4]These apparent gender mismatch effects at the disambiguating and post-disambiguating regions may be due to preceding differences in lexical material between gender match and gender mismatch sentences (Ken's mum/dad). However, it is unclear why this would influence ambiguous sentences only. As this effect is not informative of L1/L2 differences, we do not discuss it in more detail here.

https://doi.org/10.1017/S1366728921000195 Published online by Cambridge University Press

https://doi.org/10.1017/S1366728921000195 Published online by Cambridge University Press

Table 3. Reading times for three eye-movement measures at four regions of texts in Experiment 2 (SDs in parentheses).

	Disambiguating region (decided)		Spillover Region (to prepare)		Reflexive Region (himself)		2nd Spillover Region (a cold drink)	
	Native Speakers	Non-Native Speakers	Native Speakers	Non-Native Speakers	Native Speakers	Non-Native Speakers	Native Speakers	Non-Native Speakers
First pass time								
Amb, Gender Match	295 (157)	352 (179)	283 (141)	369 (183)	207 (73)	268 (133)	378 (215)	594 (333)
Amb, Gender Mismatch	323 (164)	372 (220)	276 (131)	377 (180)	241 (141)	271 (115)	387 (216)	610 (339)
Unamb, Gender Match	270 (133)	336 (164)	268 (113)	362 (192)	210 (71)	259 (110)	393 (217)	578 (334)
Unamb, Gender Mismatch	277 (133)	325 (157)	283 (123)	365 (177)	234 (104)	272 (131)	428 (274)	600 (351)
Regression path duration								
Amb, Gender Match	522 (726)	605 (654)	568 (689)	874 (1262)	332 (524)	350 (332)	834 (1236)	1115 (1396)
Amb, Gender Mismatch	625 (793)	747 (1107)	700 (927)	900 (1256)	387 (879)	440 (691)	1007 (1674)	1396 (1797)
Unamb, Gender Match	400 (440)	430 (399)	404 (460)	409 (294)	252 (215)	280 (168)	566 (555)	1116 (1324)
Unamb, Gender Mismatch	410 (546)	447 (467)	382 (413)	435 (309)	387 (671)	357 (373)	965 (1112)	1382 (1351)
Total viewing time								
Amb, Gender Match	517 (379)	746 (524)	447 (266)	695 (497)	285 (186)	438 (317)	497 (294)	882 (612)
Amb, Gender Mismatch	685 (545)	871 (651)	558 (406)	821 (636)	390 (302)	546 (410)	596 (424)	1061 (779)
Unamb, Gender Match	422 (271)	539 (422)	371 (221)	529 (339)	284 (159)	403 (269)	481 (264)	875 (549)
Unamb, Gender Mismatch	477 (369)	574 (366)	454 (333)	617 (372)	420 (324)	534 (386)	650 (477)	1007 (611)

NB: Amb = Ambiguous, Unamb = Unambiguous

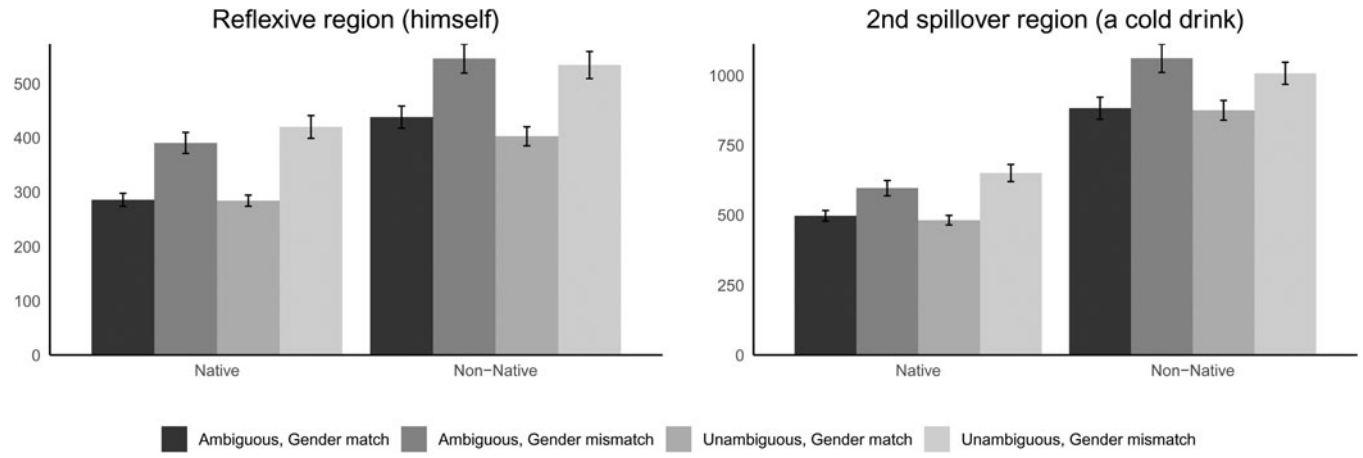


Figure 1. Total viewing times in milliseconds at the reflexive and 2nd spillover regions in Experiment 2. Error bars indicate standard errors.

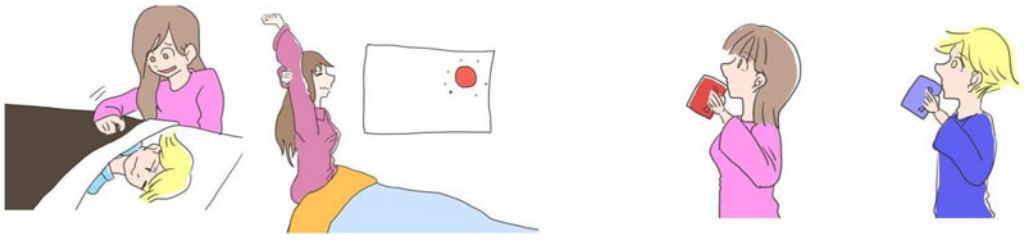


Figure 2. Example picture pairs used to test subordinate clauses (left two pictures) and main clauses (right two pictures) in Experiment 3.

In regression path duration, there was a significant interaction between gender and region, due to larger gender mismatch effects at the spillover region (*estimate* = 0.236, $t = 6.50$, $SE = 0.04$, $p < .001$) than the reflexive region (*estimate* = 0.129, $t = 4.50$, $SE = 0.03$, $p < .001$). There was also a marginal interaction between ambiguity and gender. However, pairwise comparisons showed significant gender mismatch effects in both ambiguous (*estimate* = 0.134, $t = 3.99$, $SE = 0.03$, $p < .001$) and unambiguous (*estimate* = 0.231, $t = 5.87$, $SE = 0.04$, $p < .001$) conditions.

Total viewing times also showed a significant interaction between gender and region due to different sized gender mismatch effects at the two regions (reflexive region, *estimate* = 0.247, $t = 7.71$, $SE = 0.03$, $p < .001$; spillover region *estimate* = 0.173, $t = 6.95$, $SE = 0.03$, $p < .001$).

Discussion

Consistent with previous studies, Experiment 2 showed that both L1 and L2 participants encountered reading difficulty upon disambiguation (Frazier & Rayner, 1982; Juffs & Harrington, 1996). However, we did not find significant evidence of smaller garden-path effects for L2ers (Jacob & Felser, 2016). Regarding reanalysis, there was evidence of gender mismatch effects, irrespective of ambiguity, at the reflexive and second spillover regions that were not significantly modulated by group. This suggests that both L1ers and L2ers constructed the correct syntactic structure after reanalysis, at least to the extent that Binding Principle A was at play, which replicates the L1 findings from Slattery et al. (2013).

Experiment 2 suggests that L2ers conduct syntactic reanalysis during online reading. This indicates that L2 reanalysis difficulty cannot be accounted for entirely by the incomplete reanalysis hypothesis. Experiments 3/4 further explored L2 reanalysis. Experiment 3 replicated Experiment 1 using a different task, while Experiment 4 tested how misinterpretations linger during online processing.

Experiment 3

Experiment 3 used sentence-picture matching. Participants read temporarily ambiguous (9a) and unambiguous (9b) sentences, and were then shown one of two pictures pairs (Figure 2). The subordinate clause picture pair denotes either the correct or incorrect action of the subordinate clause ("the lady woke up/the lady woke up her husband"), while the main clause picture pair depicts either the correct or incorrect action of the main clause ("her husband drank some coffee/the lady drank some coffee"). Participants chose which picture they thought best corresponded to the sentence. Our predictions were the same as in Experiment 1.

(9a) Ambiguous

After the lady woke up her husband in the apartment drank some coffee.

(9b) *Unambiguous*

After the lady woke up, her husband in the apartment drank some coffee.

https://doi.org/10.1017/S1366728921000195 Published online by Cambridge University Press

Method

Participants

Forty L1 English speakers (7 males, mean age 19; range 18–23) and 40 L2 English speakers (14 males, mean age 23; range 18–47) of various L1 backgrounds,[5] none of whom took part in Experiments 1/2, from the University of Reading community participated in Experiment 3. Participants received course credit or payment. L2 participants started learning English from age eight onwards (mean age of onset 8.9; SD 1.1; range 8–11) and also completed the OPT, which showed that they were upper intermediate-advanced English language learners (mean 76; SD 10.6; range 51–94).

Materials

Experiment 3 employed 24 sentences as in (9), using only RAT or reciprocal verbs, and four coloured pictures for each experimental set. Two of the pictures tapped the interpretation of the subordinate clause, while the other two examined the interpretation of the main clause. The experiment additionally included 84 fillers of a variety of different constructions, accompanied by two pictures. The experimental and filler sentences were presented in a counterbalanced Latin square design.

Procedure and data analysis

The procedure was identical to Experiment 1, except that after each sentence, participants saw one picture pair, and chose which picture that they felt best matched the content of the sentence. The experiment was administered using the Ibex Farm web-based platform (Drummond, 2013) but was completed by participants in a traditional lab setting.

Results

Descriptive and inferential statistics are encapsulated in Tables 1 and 2. Accuracy rates to fillers were 94% for L1 participants (range 69–100) and 95% for L2 participants (range 83–100).

There was a significant main effect of question type, with lower accuracy rates for subordinate clause than main clause questions. There was also a significant main effect of ambiguity, with lower accuracy for ambiguous sentences, qualified by a marginal two-way interaction between ambiguity and group. This suggested significantly lower accuracy rates for L2 than L1 participants in the ambiguous conditions only (ambiguous: *estimate* = 0.960, z = 2.52, SE = 0.38, p = .012; unambiguous: *estimate* = 0.155, z = 0.30, SE = 0.52, p = .765).

Discussion

As in Experiment 1, Experiment 3 indicated lingering misinterpretation in L1ers and L2ers. There was also some evidence that L2ers had particular difficulty with reanalysis, although the relevant interaction between group and ambiguity was only marginally significant. As mentioned for Experiment 1, the finding that both groups chose pictures incorrectly some of the time in the ambiguous main clause condition may not be expected if readers reanalyse the temporarily ambiguous noun phrase as the main clause subject (Slattery et al., 2013). We return to these issues in the General Discussion, but first report Experiment 4, which examined lingering misinterpretation during online processing.

[5]First languages were Greek (8), Bulgarian (5), Italian (5), French (3), Indonesian (3), Polish (3), Lithuanian (2), Spanish (2), Turkish (2), Arabic (1), Bangladeshi (1), Chinese (1), Dutch (1), Finnish (1), German (1), and Thai (1).

Experiment 4

Experiment 4 examines whether misinterpretation influences subsequent sentence processing. Slattery et al. (2013) used an eye-tracking paradigm whereby a continuation sentence following their critical sentences was always consistent with the correct interpretation of the temporarily ambiguous sentences. We extended their design by including not only consistent but also inconsistent sentences as in (10).

(10a) *Ambiguous, Consistent Continuation*

When the mother dressed her son at home called the dog.
It was clear that the mother was dressing herself formally for an important ceremony.

(10b) *Unambiguous, Consistent Continuation*

When the mother dressed, her son at home called the dog.
It was clear that the mother was dressing herself formally for an important ceremony.

(10c) *Ambiguous, Inconsistent Continuation*

When the mother dressed her son at home called the dog.
It was clear that the mother was dressing her son formally for an important ceremony.

(10d) *Unambiguous, Inconsistent Continuation*

When the mother dressed, her son at home called the dog.
It was clear that the mother was dressing her son formally for an important ceremony.

It was tiring.

The first sentence is either temporarily ambiguous (10a/10c) or unambiguous (10b/10d). The subordinate clause used either a RAT or reciprocal verb for the same reason as Experiments 1/3. In (10a/b), the continuation sentence refers to the correct interpretation of the first sentence ("the mother was dressing herself") and is thus consistent with it. The continuation sentence in (10c/d) on the other hand refers to the initial misinterpretation ("the mother was dressing her son") and thus is inconsistent with the correct interpretation of the first sentence.

Importantly, this (in)consistency may be reversed or attenuated depending on how the first sentence is interpreted and how strongly misinterpretation lingers. If the correct analysis of the first sentence is constructed, a main effect of consistency is expected in the second sentence, with longer reading times in inconsistent (10c/d) than consistent (10a/b), irrespective of ambiguity. However, if the initial misinterpretation ("the mother dressed her son") lingers, inconsistency effects should be modulated by ambiguity. In the inconsistent conditions, reading times of the continuation sentence should be shorter for ambiguous (10c) than unambiguous sentences (10d), as a result of the continuation sentence in (10c) being misinterpreted as being consistent with the ambiguous first sentence. For the consistent conditions, if the misinterpretation ("the mother dressed her

https://doi.org/10.1017/S1366728921000195 Published online by Cambridge University Press

Table 4. Reading times for three eye-movement measures at three regions of texts in Experiment 4 (SDs in parentheses).

	Disambiguating region (called)		Spillover Region (the dog)		Critical Region (dressing her son/herself)		2nd Spillover Region (formally for an)	
	Native Speakers	Non-Native Speakers	Native Speakers	Non-Native Speakers	Native Speakers	Non-Native Speakers	Native Speakers	Non-Native Speakers
First pass time								
Amb, Consistent	296 (150)	286 (130)	362 (226)	475 (256)	237 (106)	293 (132)	401 (232)	540 (359)
Unamb, Consistent	271 (119)	282 (130)	424 (269)	530 (285)	268 (159)	287 (138)	409 (269)	506 (305)
Amb, Inconsistent	293 (142)	316 (162)	363 (213)	500 (294)	288 (127)	373 (220)	420 (247)	498 (298)
Unamb, Inconsistent	284 (142)	281 (132)	421 (268)	550 (288)	310 (147)	384 (214)	405 (260)	496 (325)
Regression path duration								
Amb, Consistent	602 (753)	393 (360)	1763 (1778)	1753 (1748)	317 (612)	408 (687)	552 (512)	704 (617)
Unamb, Consistent	363 (354)	377 (398)	887 (859)	1093 (1155)	308 (222)	381 (288)	540 (530)	617 (439)
Amb, Inconsistent	624 (754)	475 (466)	1682 (1669)	1696 (1708)	344 (196)	439 (436)	625 (688)	691 (826)
Unamb, Inconsistent	389 (401)	339 (288)	835 (858)	1065 (1023)	429 (615)	484 (548)	701 (838)	699 (479)
Total viewing time								
Amb, Consistent	657 (399)	771 (615)	745 (465)	1160 (773)	336 (216)	457 (291)	639 (453)	923 (646)
Unamb, Consistent	457 (308)	563 (342)	695 (451)	1021 (654)	376 (253)	477 (305)	643 (444)	885 (580)
Amb, Inconsistent	721 (466)	835 (611)	748 (455)	1178 (800)	453 (332)	622 (458)	705 (461)	902 (614)
Unamb, Inconsistent	513 (402)	579 (390)	761 (569)	1068 (622)	539 (376)	722 (492)	732 (441)	1001 (661)

NB: Amb = Ambiguous, Unamb = Unambiguous

son") lingers, the ambiguous consistent continuation condition should be misperceived as being inconsistent. This would lead to longer reading times in (10a) than (10b). Thus, the crucial prediction is whether a main effect of consistency is observed, or an interaction between consistency and ambiguity.

Regarding L1/L2 differences, the initial misinterpretation may persist more strongly in L2ers than L1ers (Cunnings, 2017). In this case, lingering effects in consistent or inconsistent sentences may be larger for L2ers than L1ers.

Methods

Participants

The participants in Experiment 4 were identical to those in Experiment 3. Experiment 4 was conducted at least one week before Experiment 3.

Materials

Experimental materials comprised 24 sets of texts like (10) in a Latin square design with two levels of ambiguity (ambiguous/unambiguous) and consistency (consistent/inconsistent). Each text contained a temporarily ambiguous or unambiguous sentence on the first line, a critical continuation sentence across two lines and a wrap-up sentence at the end of the second line. For continuation sentences, the line-break appeared after the complementizer ("that"). The experiment also contained 72 filler texts, which always took up two lines on the screen. All experimental texts and two-thirds of the fillers were followed by a yes/no comprehension question that did not tap any of the critical manipulations.

Procedure and data analysis

The procedure was the same as Experiment 2. L2 participants completed a vocabulary test that tested their knowledge of the words used for the subordinate clause subject and verb, and the main clause subject.

The calculation of eye-tracking measures, data exclusion criteria and data analysis procedure were identical to Experiment 2, except that a fixed effect of consistency (consistent/inconsistent) was included instead of gender. To test for garden-path effects, we specified the main clause verb of the first sentence as the disambiguating region ("called") and the rest of the sentence as the first spillover region ("the dog"). To test for consistency effects, the critical region was defined as the text that denoted the (in)consistency effect ("her son/herself") and the second spillover region ("formally for an") was defined as the rest of the sentence except the last two words to avoid wrap-up effects. Skipping rates were 8% for the L1 data and 3% for the L2 data across all regions. Trials including words that L2 participants did not know were removed from analysis, which affected less than 0.1% of the L2 data.

Results

Overall accuracy rates of comprehension questions were 89% for L1 participants (range 75–97%) and 88% for L2 participants (range 76–97%). The reading time data are presented in Table 4. A summary of the inferential statistics can be found in Online Supplement 2.

Disambiguating and 1st spillover regions

There was a significant main effect of ambiguity that was qualified by a significant ambiguity by region interaction in first pass times. At the disambiguating region, first pass times were marginally longer for ambiguous than unambiguous sentences (*estimate* = 0.044, $t = 1.97$, $SE = 0.02$, $p = .058$), but this pattern was reversed

https://doi.org/10.1017/S1366728921000195 Published online by Cambridge University Press

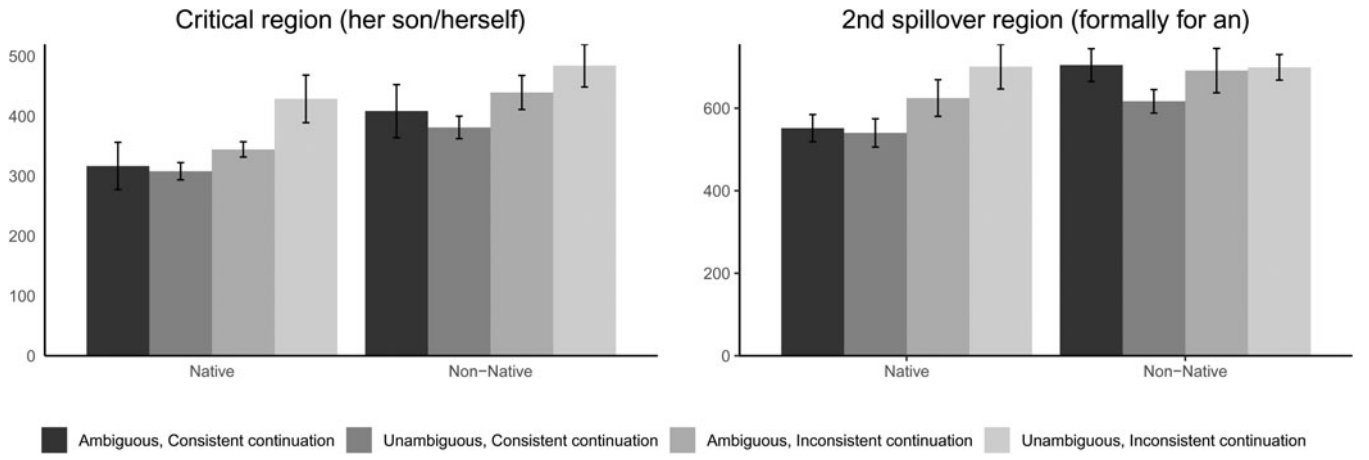


Figure 3. Regression path duration in milliseconds at the critical and 2nd spillover regions in Experiment 4. Error bars indicate standard errors.

at the spillover region (*estimate* = 0.128, $t = 4.36$, $SE = 0.03$, $p < .001$).

While this effect at the spillover region may appear counter-intuitive, shorter first pass times for ambiguous conditions may occur if readers quickly regressed out of this region (e.g., Sturt, 2007). Indeed, consistent with this interpretation of first pass times, regression path times indicated a significant main effect of ambiguity, with longer reading times for ambiguous sentences. This was qualified by a significant ambiguity by region interaction, with longer reading times for ambiguous than unambiguous sentences at both disambiguating (*estimate* = 0.202, $t = 6.22$, $SE = 0.03$, $p < .001$) and spillover (*estimate* = 0.471, $t = 8.88$, $SE = 0.05$, $p < .001$) regions, though the effect was larger at the spillover region. There was also a two-way interaction between ambiguity and group. Pairwise comparisons showed garden paths for both groups but with a larger effect in the L1ers (L1: *estimate* = 0.419, $t = 9.90$, $SE = 0.04$, $p < .001$; L2: *estimate* = 0.265, $t = 6.02$, $SE = 0.04$, $p < .001$).

In total viewing times, there was a significant main effect of ambiguity, an ambiguity by region interaction and a significant three-way interaction between ambiguity, group and region. However, 2×2 analysis by region showed only significant and marginal main effects of ambiguity due to garden-path effects at the disambiguating (*estimate* = 0.315, $t = 8.10$, $SE = 0.04$, $p < .001$) and spillover (*estimate* = 0.068, $t = 1.97$, $SE = 0.03$, $p = .058$) regions.

Critical and 2nd spillover regions

At the critical and second spillover regions in the continuation sentence, there was a significant main effect of consistency in all measures, with longer reading times in inconsistent conditions.

In first pass reading times, there was a marginal interaction between consistency, group and region. Planned comparisons by region revealed a significant main effect of consistency only in the critical region due to longer reading times for inconsistent than consistent sentences (critical region: *estimate* = 0.212, $t = 6.23$, $SE = 0.03$, $p < .001$; second spillover region: *estimate* = 0.011, $t = 0.38$, $SE = 0.03$, $p = .705$). There was also a significant interaction between ambiguity and region. Pairwise comparisons showed a significant main effect of ambiguity only in the critical region due to shorter reading times for ambiguous than unambiguous sentences (critical region: *estimate* = 0.041, $t = 2.27$, $SE = 0.02$, $p = .023$; spillover region: *estimate* = 0.024, $t = 0.78$, $SE = 0.03$, $p = .433$).

Regression path duration showed a significant interaction between ambiguity and consistency. Pairwise comparisons indicated that for inconsistent sentences, reading times were significantly shorter for ambiguous than unambiguous sentences (*estimate* = 0.087, $t = 3.40$, $SE = 0.03$, $p < .001$), showing lingering misinterpretation. Reading times did not differ in consistent sentences (*estimate* = 0.002, $t = 0.06$, $SE = 0.03$, $p = .955$). This effect of lingering misinterpretation is illustrated in Figure 3. There was also a consistency by region interaction, as reading times were longer for inconsistent than consistent sentences only in the critical region (critical region: *estimate* = 0.211, $t = 6.25$, $SE = 0.03$, $p < .001$; second spillover region: *estimate* = 0.074, $t = 1.63$, $SE = 0.05$, $p = .118$).

Total viewing times showed a significant interaction between consistency and region due to larger inconsistency effects at the critical region (*estimate* = 0.317, $t = 8.30$, $SE = 0.04$, $p < .001$) than the second spillover region (*estimate* = 0.106, $t = 2.39$, $SE = 0.04$, $p = .024$). There was also a significant interaction between ambiguity and region. Planned comparisons showed a significant main effect of ambiguity only at the critical region due to reduced reading times for ambiguous sentences (critical region: *estimate* = 0.118, $t = 440$, $SE = 0.03$, $p < .001$; second spillover region: *estimate* = 0.045, $t = 1.49$, $SE = 0.03$, $p = .144$).

Discussion

As in Experiment 2, Experiment 4 showed that L1 and L2 participants had difficulty reading ambiguous sentences due to reanalysis. One measure (namely, regression path duration) also indicated smaller garden-path effects for L2ers than L1ers (Jacob & Felser, 2016).

In the continuation sentence, longer reading times for inconsistent conditions suggest both groups generally conducted syntactic reanalysis. Importantly, there was also evidence of lingering misinterpretation. This was most evident in regression path duration, which was significantly shorter following ambiguous than unambiguous sentences in the inconsistent conditions. This suggests the initially assigned misinterpretation persisted to the extent that it influenced reading of the continuation sentence. Although this is compatible with misinterpretation lingering in

https://doi.org/10.1017/S1366728921000195 Published online by Cambridge University Press

memory, we did not fully replicate the findings of Experiment 2 reported in Slattery et al., (2013), who did not test inconsistent sentences but did report lingering effects in consistent sentences. We, however, observed differences between inconsistent but not consistent conditions. Additionally, we did not find significant differences in terms of lingering misinterpretation between L1ers and L2ers. We discuss these results, along with our other findings, in more detail below.

General Discussion

The aims of the present study were to investigate whether L2ers have more difficulty in reanalysis than L1ers and why L2 reanalysis difficulty occurs. Experiments 1/3 suggested that L2ers are more persistent with initial misanalyses than L1ers. Experiment 2 provided evidence that both L1ers and L2ers conduct syntactic reanalysis, and Experiment 4 showed that misinterpretation persists past syntactic disambiguation and influences subsequent sentence processing. Below, the implications of these results are discussed.

L2 reanalysis processes

Experiments 1/3 provide some support for the claim that reanalysis is more difficult for L2ers (e.g., Jacob & Felser, 2016; Pozzan & Trueswell, 2016). In Experiment 1, we found significantly lower accuracy for L2ers than L1ers in ambiguous sentences, but only in main clause questions. In Experiment 3, L2ers tended to have lower comprehension accuracy than L1ers for ambiguous sentences, for both question types. As such, the direction of effects is compatible with previous studies indicating increased reanalysis difficulty in ambiguous sentences for L2ers.

We considered two accounts of this L2 reanalysis difficulty. If L2ers do not conduct syntactic reanalysis, we reasoned that L2ers should have lower accuracy following ambiguous sentences than L1ers, irrespective of the question type. Alternatively, if L2ers conduct syntactic reanalysis but the initial misinterpretation lingers, L2ers should have lower accuracy than L1ers for subordinate clause questions only. Although our offline tasks suggested increased reanalysis difficulty in L2ers, the pattern of results did not consistently provide evidence either way. On the other hand, the results of our online experiments provided clearer evidence in this regard. L1ers and L2ers both showed clear gender mismatch effects in Experiment 2, which suggests the ambiguous noun phrase was syntactically reanalysed as the main clause subject. Indeed, we did not find significant evidence to suggest that L2ers conducted syntactic reanalysis any less successfully than L1ers in this experiment. The effects of ambiguity on sentence continuations in Experiment 4 suggested lingering misinterpretation in both L1ers and L2ers. Although we did not find evidence of increased lingering misinterpretation in L2ers than L1ers in Experiment 4, as predicted by Cunnings (2017), these results are consistent with the idea that L1ers and L2ers conduct syntactic reanalysis to a similar degree but have difficulty erasing the initial misinterpretation from memory.

Additionally, both online experiments showed clear garden-path effects in L2ers. Jacob and Felser (2016) reported larger garden-path effects for L1ers than L2ers, which they took to indicate that L2ers do not conduct reanalysis as consistently as L1ers. Although we found a similar pattern in regression path times at the disambiguating region in Experiment 4, this effect was not found in other measures in this experiment, nor in any measure in Experiment 2. As such, we did not consistently find evidence in the size of garden-path effects to suggest that L2ers initiated/completed reanalysis less often than L1ers during processing.

We acknowledge that individual differences such as properties of the L2ers' L1, their English proficiency and the age of English onset may have affected our results, either in terms of the size of garden-path effects or persistence of lingering misinterpretations. We tested L2ers with a variety of L1 backgrounds, as previous research has shown garden-path effects in L2ers irrespective of L1 background (e.g., Juffs, 2004). However, as illustrated by the standard deviations in Table 1, comprehension accuracy rates following ambiguous sentences are widely distributed for L2ers, with some achieving comprehension rates comparable to L1ers. Given some studies show that individual differences influence the size of garden-path effects (Jegerski, 2012; Havik, Roberts, van Hout, Schreuder & Haverkort, 2009; Hopp, 2015), how individual differences may influence lingering misinterpretation may be one key to clarifying L1/L2 differences.[6]

Good enough language processing in L1 and L2 comprehension

Experiments 1, 3 and 4 showed that initial misinterpretations linger in both L1ers and L2ers at the offline and online levels. This lingering effect is compatible with good-enough processing, which predicts that comprehenders do not always erase previously created representations that turn out to be incorrect (e.g., Ferreira et al., 2001). Slattery et al. (2013) claimed L1ers conduct syntactic reanalysis and argued that lingering effects result from co-existing representations of the initially assigned and globally correct interpretations. Experiment 2 replicated this finding and extended it to L2ers.

One finding from Experiments 1 and 3, which the good-enough account might not predict, is that participants sometimes answered ambiguous main clause questions incorrectly. If the temporarily ambiguous noun phrase is successfully reanalysed as the main clause subject, accuracy to ambiguous main clause questions should be as high as those to unambiguous ones. In their Experiment 2, Christianson et al. (2001) tested interpretation of both the subordinate clause and the main clause in a similar way to our offline experiments. They reported a significant interaction between question type and ambiguity. Although errors occurred following ambiguous sentences for questions tapping both clauses, they were much more frequent for subordinate clauses (62%) than main clauses (12%), compatible with the claim that the temporarily ambiguous noun phrase is reanalysed but the initial misinterpretation lingers. However, across our offline experiments, the numerical differences between error rates for subordinate clause questions (L1ers 37%, L2ers 47%) and main clause questions (L1ers 22%, L2ers 38%) were smaller. One difference between our study and Christianson et al. is how questions were asked. While we asked wh-questions (e.g., "Who laughed very loudly?"; Experiment 1) or used sentence-picture matching (Experiment 3), Christianson et al.'s questions always asked yes/no questions (equivalent to "Did the baby laugh?" for our example (6)) that always referred to the correct interpretation. This may have biased participants in Christianson et al. towards the correct interpretation (Tabor et al., 2004) more often than in our study.

[6]We tested whether L2 proficiency affected our results by fitting (generalised) linear mixed-effect models with centred OPT scores as a continuous predictor. However, these analyses did not provide consistent evidence that higher proficiency L2ers behaved more nativelike in terms of reanalysis across our experiments.

https://doi.org/10.1017/S1366728921000195 Published online by Cambridge University Press

The comparatively lower accuracy to main clause questions in our offline experiments suggests that in sentences like "After the mother dressed the baby in the living room laughed very happily", readers sometimes misinterpreted "the mother" to be the subject of "laughed", at least during the post-sentence phase of our offline tasks. One potential account of this might be that readers picked the wrong answer here because the subordinate clause subject ("the mother") is more unambiguously a subject than the main clause subject ("the baby"), which is initially misinterpreted as an object.

While our offline results for main clause questions may thus suggest reanalysis of the ambiguous noun phrase is not always conducted, the results from Experiment 2, like Slattery et al. (2013), found no evidence to suggest syntactic reanalysis is not conducted during online reading. One potential account of this discrepancy is that the reflexive in Experiment 2 (and Slattery et al.), in referring to the main clause subject soon after disambiguation, may have reinforced the correct interpretation of the temporarily ambiguous noun, compared to our offline experiments that did not include a reflexive. How such effects may influence lingering misinterpretation in L1ers and L2ers, either in offline experiments like Experiments 1/3 or online tasks like Experiments 2/4, may be a fruitful avenue of future research.

Finally, how L1ers and L2ers are prone to misinterpretation during sentence processing needs further exploration. This issue derives from the different results between Experiment 2 of Slattery et al. (2013) and Experiment 4 of the present study. Specifically, while Slattery et al. showed lingering misinterpretation when reading subsequent text that was CONSISTENT with the globally correct analysis of the temporary ambiguity, the present study showed such effects only when the subsequent text contained INCONSISTENT information. Although both effects are consistent with lingering misinterpretation, further research examining the relative size of these two effects is required. More generally, although our results are broadly consistent with "good-enough" processing, these inconsistencies with previous L1 studies highlight the need for increased replication in psycholinguistics (Vasishth, Mertzen, Jäger & Gelman, 2018), in both L1 and L2 processing.

Conclusion

The present study reported four experiments investigating L1 and L2 reanalysis. Two offline experiments were consistent with L2ers having more difficulty reanalysing garden-path sentences than L1ers. The two eye-movement experiments showed that L1ers and L2ers conduct syntactic reanalysis during online reading but that the initially assigned misinterpretation is not completely discarded from memory after reanalysis. Taken together, we argue that our results suggest that L2 reanalysis difficulty at least partly results from a failure to erase initially assigned misinterpretations from memory, rather than an inability to conduct syntactic reanalysis.

Supplementary Material. For supplementary material accompanying this paper, visit http://doi.org/10.1017/S1366728921000195. *Online Supplement 1*, containing the experimental items from Experiments 1–4, and *Online Supplement 2*, containing the inferential statistics for Experiments 2 and 4, are available as a single file.

Acknowledgments. We would like to thank the editor and three reviewers for comments on previous drafts of this paper. The study reported here was supported by a Language Learning Dissertation Grant awarded to HF.

Competing interests. The authors have no competing interests to declare.

References

Allan D (2004) *The Oxford placement test*. Oxford, UK: Oxford University Press.

Baayen RH, Davidson DJ and Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* **59**, 390–412.

Barr D, Levy R, Scheepers C and Tily H (2013) Random-effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language* **68**, 255–278.

Chomsky N (1981) *Lectures on government and binding*. Dordrecht: Foris.

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

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

Clahsen H and Felser C (2017) Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition* **40**, 693–706.

Cunnings I (2017) Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition* **20**, 659–678.

Cunnings I and Sturt P (2018) Coargumenthood and the processing of pronouns, *Language, Cognition and Neuroscience* **33**, 1235–1251.

Drummond A (2013) Ibex Farm. Available online, http://spellout.net/ibexfarm/.

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

Ferreira F, Christianson K and Hollingworth A (2001) Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research* **30**, 3–20.

Ferreira F and McClure KK (1997) Parsing of garden-path sentences with reciprocal verbs. *Language and Cognitive Processes* **12**, 273–306.

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

Gerth S, Otto C, Nam Y and Felser C (2017) Strength of garden-path effects in native and non-native speakers' processing of subject-object ambiguities. *International Journal of Bilingualism* **21**, 125–144.

Gibson E (1991) *A computational theory of human linguistic processing: Memory limitations and processing breakdown* (Unpublished doctoral thesis) School of Computer Science, Carnegie Mellon University, PA.

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

Hopp H (2015) Individual differences in the second language processing of object–subject ambiguities. *Applied Psycholinguistics* **36**, 129–173.

Jacob G and Felser C (2016) Reanalysis and semantic persistence in native and non-native garden-path recovery. *The Quarterly Journal of Experimental Psychology* **69**, 907–925.

Jegerski J (2012) The processing of temporary subject– object ambiguities in native and near-native Mexican Spanish. *Bilingualism: Language and Cognition* **15**, 721–735.

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

Juffs A and Harrington M (1996) Garden path sentences and error data in second language processing research. *Language Learning* **46**, 283–323.

Kaschak MP and Glenberg AM (2004) This construction needs learned. *Journal of Experimental Psychology: General* **133**, 450–467.

Kuznetsova A, Brockhoff P and Christensen R (2017) lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* **82**, 1–26.

MacDonald MC, Pearlmutter NJ and Seidenberg MS (1994) The lexical nature of syntactic ambiguity resolution. *Psychological Review* **101**, 676–703.

Malyutina S and den Ouden DB (2016) What is it that lingers? Garden-path (mis)interpretations in younger and older adults. *The Quarterly Journal of Experimental Psychology* **69**, 880–906.

Pozzan L and Trueswell J (2016) Second language processing and revision of garden-path sentences: A visual world study. *Bilingualism: Language and Cognition* **19**, 636–643.

https://doi.org/10.1017/S1366728921000195 Published online by Cambridge University Press

R Core Team (2018) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rohde D (2010) *Linger*. Available online, https://tedlab.mit.edu/~dr/Linger/

Slattery TJ, Sturt P, Christianson K, Yoshida M and Ferreira F (2013) Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language* **69**, 104–120.

Sturt P (2003) The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language* **48**, 542–562.

Sturt P (2007) Semantic re-interpretation and garden-path recovery. *Cognition* **105**, 477–488.

Sturt P, Pickering MJ and Crocker MW (1999) Structural change and reanalysis difficulty. *Journal of Memory and Language* **40**, 136–150.

Tabor W, Galantucci B and Richardson D (2004) Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language* **50**, 355–370.

Van Gompel RPG, Pickering MJ and Traxler MJ (2000) Unrestricted race: A new model of syntactic ambiguity resolution. In Kennedy A, Radach R, Heller D and J. Pynte (Eds.), *Reading as a perceptual process* Oxford: Elsevier 621–648.

van Gompel RPG, Pickering MJ, Pearson J and Jacob G (2006) The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language* **55**, 335–362.

Vasishth S, Mertzen D, Jäger LA and Gelman A (2018) The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language* **103**, 151–175.

von der Malsburg T and Angele B (2017) False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language* **94**, 119–133.

https://doi.org/10.1017/S1366728921000195 Published online by Cambridge University Press