Sensitivity to event structure in passives supports deep processing in L1 and L2

Katrina Geraghty, Nino Grillo, Shayne Sloggett

University of York

Abstract

A key question in second language research is whether native (L1) and non-native (L2) sentence processing are fundamentally different. Recent L1 processing research has questioned the long-held assumption that passives are harder to process than actives: passive complexity appears to be determined by event structure (Paolazzi, Grillo, Alexiadou, & Santi, 2019; Paolazzi et al., 2021). We replicate these results with a different method (maze task); only passives of states appear to be more difficult to process than actives, inconsistent with a good-enough account. We also present evidence that L2 learners are capable of recruiting similarly nuanced processing mechanisms in understanding passives. L2 learners display the same interaction of event structure and passivization. Taken together, the results appear inconsistent with shallow processing accounts of both L1 and L2 processing.

Introduction

Passive sentences have long been assumed to be more difficult to parse and interpret than active sentences. As discussed in Paolazzi et al. (2021), this assumption, largely based on offline results showing greater comprehension errors in passives than actives (Ferreira, 2003; Street & Dąbrowska 2010), motivated two families of accounts: the Syntactic Complexity account (e.g. Borer & Wexler 1987; Osterhout & Swinney 1993), and the Good-Enough account (Christianson, Hollingworth, Halliwell & Ferreira, 2001; Ferreira, Bailey & Ferraro, 2002; Ferreira, 2003; Ferreira & Christianson, 2016; Ferreira & Patson, 2007; Karimi & Ferreira, 2016).

Syntactic complexity accounts attribute the higher complexity of passives to the more complex derivation of passives, which involve, among other things, movement of the internal argument (e.g. *the man*) in (1) from the position where it receives a thematic interpretation as the patient of *promote* to the subject position.

(1) [The man]$_i$ was promoted [e]$_i$ by the woman.

The additional syntactic complexity resulting from the noun phrase (NP) movement distinguishes passives from active sentences, with the former considered inherently more costly to processing resources, and subsequently more difficult to interpret. The alleged higher syntactic complexity of passives is also an argument for the Good-Enough theory. Consider sentence (2) (Ferreira et al., 2002:13):

(2) The dog was bitten by the man.

Building on observations of lingering misinterpretations of garden-path sentences, Ferreira et al. (2002) suggest that comprehenders often misinterpret the meaning of passive constructions like (2) due to semantic heuristics overriding bottom-up cues to syntactic structure: the atypical patient-agent order makes them harder to parse and interpret, leading comprehenders to an interpretation that is 'good-enough'. Under this approach, there are two main routes for sentence processing: algorithmic and heuristic, though the interaction between these two processing streams remains a point of variation in specific models. Algorithmic processing represents deep, structural analysis, and is assumed to be more costly and time-consuming to the parser. The heuristic processing stream is considered a faster processing strategy focussing on lexical-semantic and pragmatic information. The Good-Enough approach proposes that whilst the parser simultaneously draws on algorithmic and heuristic information, sentence representations need not always be fully specified. That is, comprehenders may occasionally rely on faster semantic and lexically-grounded heuristics such as the NVN (noun, verb, noun) strategy, in which the first noun is interpreted as the agent, and the second noun as the patient or theme of the verb. In passives, slower, algorithmic parsing processes are required to correct the initial 'agent-first' misinterpretation. Supporting evidence from tasks specifically targeting thematic roles in passive and active sentences suggest that passives elicit a higher error rate than actives (Ferreira, 2003; Christianson, Luke & Ferreira, 2010).

In sum, the Good-Enough approach predicts that passives should be mis-interpreted precisely because they are generally costly to process: comprehenders may be more reliant on heuristics when the underlying syntax is complex, thus avoiding slower, algorithmic processing. The assumption that passives are harder to parse than actives is belied by online results showing that reading times are in fact *shorter* in passives than actives (Carrithers 1989; Paolazzi et al. 2019, 2021; Traxler, Corina, Morford, Hafer & Hoversten, 2013) and that these durational differences do not have a negative impact on comprehension.

Paolazzi et al. argue that neither traditional approaches can account for these results. Building on Gehrke and Grillo's (2007; 2009) observation that the availability of verbal passivization is largely dependent on event structure, Paolazzi et al. manipulated voice (active/passive) and event type (stative/eventive) in English with different online and offline techniques (self-paced reading, eye-tracking, acceptability judgments, comprehension questions targeting thematic role assignment) and showed that passives are not more difficult to parse and interpret than actives, but that the observed complexity effects in passive sentences are reducible to independent issues relating to the interaction of passivization and event structure and task biases on memory demands. We extend these results in two directions: first of all, we show that shorter reading times for passives are also observed when using a different method, the maze task, which is notable for requiring deep processing for successful completion. Then we show that similar reading patterns are observed in both L1 and L2 English speakers. These results provide additional arguments against shallow processing approaches to both L1 and L2 processing. In the remainder of this introduction, we briefly summarise recent research on the processing of passives, focusing primarily on Paolazzi et al.'s work and the role of event structure in modulating the complexity of passivization. This will serve as a background for the discussion of L2 processing and the novel experiment.

Passivization and event structure

Paolazzi et al. argue that previous experimental work on passives did not take into account the contrast between offline responses to comprehension questions and online results showing faster reading times for passives than for actives. Following Gehrke and Grillo (2007, 2009), they also point out that this literature largely overlooked the potential role played by event structure in modulating the complexity of passivization. Compare examples (3) and (4) (Paolazzi et al., 2021:139):

(3) The guitarist was rejected by the attractive and talented singer on Tuesday morning.

(4) The guitarist was admired by the attractive and talented singer throughout the tour.

Passivised eventive predicates such as (3) reliably signal the intended verbal interpretation. On the other hand, passivised stative predicates, as in (4), are initially compatible with an adjectival interpretation, and, in the presence of a specific by-phrase, must be coerced into a state resulting from a preceding event (Gehrke & Grillo, 2007; 2009). Across four self-paced reading (SPR) experiments, Paolazzi et al. (2019) manipulated voice (active/passive) and event type (stative/eventive) in English and found that passives were consistently read more quickly than actives at multiple regions of the sentence, including the main verb through the by-phrase. One possible reason for this is that, in comparison to simple active sentences, passive sentences in English have a richer morphological structure. That is, in line with surprisal models of sentence processing (Hale, 2001; Levy, 2008), morphological cues could serve to facilitate reading times by making upcoming morphological information more predictable in context. Participants were also slower to respond to comprehension questions following passives than actives, and responded less accurately to passives than actives in both the eventive and stative conditions. An effect of predicate type was observed in the comprehension accuracy results, such that accuracy to questions following a stative predicate were lower than to those following an eventive predicate, though no interaction was found between syntax and predicate type.

Paolazzi et al. (2019) argue that these results are incompatible with Good-Enough processing; recall that a Good-Enough account predicts that reduced reading times - indicative of comprehenders employing a faster, heuristics-based processing mechanism - should result in a failure to consistently assign thematic roles correctly. Thus, higher accuracy in offline comprehension should, if anything, be accompanied by the slower reading times indicative of deeper processing. Furthermore, in the final experiment, the researchers found correlations between working memory and offline measures, such that participants with a greater working memory capacity responded more quickly and with greater accuracy to the comprehension questions than those with a lower working memory capacity. Thus, Paolazzi et al. (2019) argue that the passive structure is not necessarily harder to parse but it appears to be more difficult to remember; particularly where a task specifically requires participants to recall the structure. Of note, Bader and Meng (2018) and Meng and Bader (2020) found differential performance in tasks involving canonical and non-canonical sentences, such that participants performed well when asked to assess sentence plausibility, but experienced greater difficulty with passives when asked to identify the patient and agent in sentences. Again, this asymmetrical performance is unexpected under a Good-Enough approach, as heuristics-based processing should lead to consistent failure in both offline tasks.

In a follow-up study, Paolazzi et al. (2021) investigated the relationship between event structure and syntax in passives using a more sensitive measure of reading: eye-tracking. The materials were adapted so that in both the stative and eventive conditions, the sentences ended with a simple temporal prepositional phrase (PP), as in (5) (Paolazzi et al., 2021, Appendix 1:1):

(5) The student was harassed by the young and inexperienced professor at the party.

As with the previous study, comprehension questions targeting thematic roles followed each sentence to measure offline processing. The results suggested an interaction between event type and syntax, particularly in late measures of processing difficulty. Whilst passive sentences were consistently read more quickly than active sentences, passives produced longer reading times than actives in the stative condition, while no such difference was found for eventive predicates. Participants were also less accurate responding to questions following passive sentences than active sentences. Taken together, the results from Paolazzi et al. (2019, 2021) strongly suggest that passive sentences are consistently read more quickly than active sentences, though when comprehension questions target thematic roles, passives elicit higher error rates than actives. Furthermore, passivised stative predicates elicit greater difficulty than passivized eventive predicates, suggesting that the cost of passivization is modulated by event type.

Adding to this picture, Grillo et al. (2019) investigated voice and event structure in German passives. Like English, German passives include the distinction between eventive and stative predicates. However, unlike English, German unambiguously distinguishes between event structures: passivised eventive predicates in German are distinguished by the auxiliary *werden* 'to become' as in (6) (Abbot-Smith & Behrens, 2006:999):

(6)	Der	reis	wurde	ge-koch-t
	the	rice	become-PAST-3SG	cook-PARTICIPLE

"the rice went through a process of being cooked"

Conversely, passivised stative predicates in German are introduced with the auxiliary *sein* 'to be' as in (7) (Abbot-Smith & Behrens, 2006:999):

(7) Der reis war ge-koch-t

the rice be-PAST-3SG cook-PARTICIPLE

"the rice was in a cooked state"

In a series of SPR experiments contrasting voice and event structure, Grillo et al. (2019) found that, like English, German passive sentences were read more quickly than active sentences by L1 German speakers. Additionally, offline comprehension for both active and passive sentences was high, irrespective of event structure. This potentially indicates that the distinctive structure for each event type facilitates comprehension in German, in line with expectation or surprisal models of sentence processing (Hale, 2001; Levy, 2008).

An event structure perspective offers a compelling argument for the differential processing of passives in L1 English sentence processing. In this paper we accomplish two independent, albeit related, goals. First of all, we aim to replicate the results from Paolazzi et al. (2019, 2021) with a different technique (the maze task). This is partly to further dig into the interesting asymmetry in results from SPR and eye-tracking, but more importantly because the maze is particularly appropriate to investigate Good-Enough claims. Due to its nature, the maze, in fact, forces readers to engage with deep processing of what they read, preventing comprehenders from relying on faster heuristics-based processing to successfully complete the task. Longer latencies are thus specifically predicted in the maze for more complex structures from a Good-Enough perspective, given the need to rely on algorithmic processing. The second question we ask is whether the account proposed by Paolazzi and colleagues can be extended to L2 processing. Our aim is not simply to address a gap in the literature, but is rooted in the desire to test a heuristics-based approach proposed for L2 sentence processing: The Shallow Structure Hypothesis (Clahsen & Felser, 2006a, 2006b, 2006c, 2018).

An extension of Good-Enough processing models, the Shallow Structure Hypothesis builds on observed differences between L1 and L2 morphological and sentence processing. The Shallow Structure Hypothesis proposes that whilst L1 speakers of a language are able to draw on the full range of syntactic and semantic information during sentence comprehension, L2 speakers, by contrast, are more likely to underuse syntactic information during online processing. Instead, L2 Speakers may be more inclined to engage a shallow parsing strategy that relies on lexical-semantic and pragmatic information, especially when the target structure is syntactically complex. That is, even highly proficient L2 speakers are more likely than L1 speakers to rely on heuristics-based processing, rather than algorithmic processing. In support of the Shallow Structure Hypothesis, studies of relative clause ambiguity resolution suggest that L2 speakers appear to be more strongly influenced by semantic information than L1 speakers (Felser, Roberts, Marinis & Gross, 2003; Papadopoulou & Clahsen, 2003). In the case of garden path sentences, L2 comprehenders appear to be more strongly guided by plausibility information than L1 comprehenders (Roberts & Felser, 2011). Additionally, L2 speakers seem to be committed to the initial incorrect parse for longer than L1 speakers, and are also less likely to reanalyse complex syntactic structures (Jacob & Felser, 2016), suggesting that L2 speakers rely more heavily on semantic information and underuse syntactic information. As successful comprehension of passives requires slower, algorithmic parsing processes to correct the initial 'agent-first' misinterpretation, comparing L1 and L2 speakers in the present study provides the opportunity to investigate whether L2 speakers are more inclined than L1 speakers to rely on the heuristic processing stream. In such a case, we would expect to see a greater comprehension penalty to passives in L2 speakers in comparison to L1 speakers.

Whilst the Shallow Structure Hypothesis is sceptical of the role of L1 influence in L2 parsing for highly proficient speakers, we would be remiss in not considering possible transfer effects relating to event semantics in the present study. Should L2 participants demonstrate sensitivity to event structure, this may suggest that L2 speakers process event structures similarly to L1 speakers. However, as stative and eventive predicates are routinely distinguished in German, this may have a facilitative effect on how L1 German speakers process passives in English. In such a case, we would expect the comprehension penalty for passivised stative predicates to be even bigger for German speakers than for English speakers, as stative predicates in English require additional coercion into a state resulting from a preceding event.

Experiment

The present study is a partial replication of experiment 3 of Paolazzi et al. (2019), a SPR experiment with a 2X2 design, contrasting syntax (active/passive) and predicate type (eventive/stative), using the experimental sentences from Paolazzi et al. (2021). As participants in Paolazzi et al.'s 2019 study were L1 speakers of English, it was necessary to consider whether the original materials were suitable for L2 speakers. Of particular concern was that the experimental items contained a complex PP, as in example (8) (Paolazzi et al., 2019, Appendix 9:14):

(8) The student was harassed by the misogynistic and racist professor in the laboratory next to the language department..

An additional concern was that this long PP construction may have introduced another intervening variable: working memory. As Paolazzi et al. (2019) reported correlations between working memory and performance in offline measures, it was necessary to minimise demands on working memory in the present study. As a result, the materials from Paolazzi et al. (2021) containing simplified PPs were used, as in Table 1. These materials were previously normed by Paolazzi et al. for plausibility.

Table 1

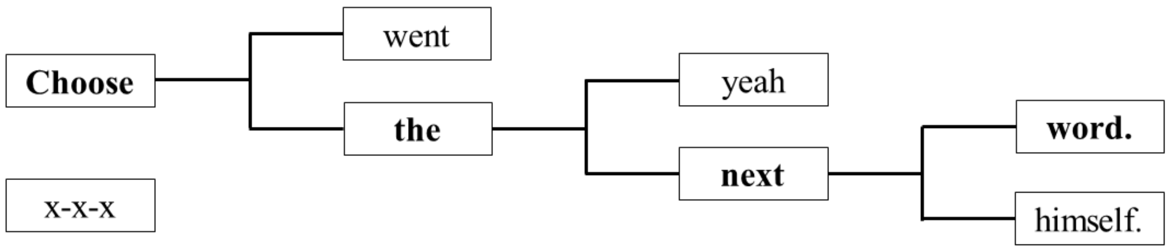
Example of experiment stimuli.

Predicate	**Syntax**	**Context Sentence**	**Sentence**	**Question**
Eventive	*Active*	Skill is essential in the hospitality industry.	The caterer paid the creative and competent chef in the afternoon.	Did the caterer pay the chef?
	Passive		The caterer was paid by the creative and competent chef in the afternoon.	Did the chef pay the caterer?
Stative	*Active*		The caterer admired the creative and competent chef since the banquet.	Did the caterer admire the chef?
	Passive		The caterer was admired by the creative and competent chef since the banquet.	Did the chef admire the caterer?

Participants read sentences with both eventive and stative predicates in the active and passive voice in a maze-task format (Witzel, Witzel & Forster, 2011; Boyce, Futrell, & Levy, 2020). A variant of SPR, the maze task involves a choice between two words at each point in the sentence, only one of which represents a licit continuation, as in Figure 1. Participants must choose the correct word to move forward, and the trial terminates if the grammatically incorrect 'foil' word is selected.

Figure 1

Example of the maze task. Bold typeface indicates the correct sentence continuation.



As highlighted by Witzel, Witzel & Forster (2011), unlike traditional SPR experiments, the maze task compels comprehenders to process input incrementally, meaning that reading times offer a more accurate measure of online sentence comprehension. Of particular interest to the present study, the maze task proscribes the use of shallow processing strategies. Specifically concerning passives, a traditional SPR experiment allows comprehenders to engage in either faster, heuristics-based processing or slower, algorithmic-based processing strategies. In the maze task, however, a Good Enough processing approach would predict slower reading times for passives, as deeper analysis must always occur in order to complete the task: failure to select the correct sentence continuation results in the trial ending. Thus, the maze task forces a degree of attention that is incompatible with shallower processing strategies. With this in mind, if L2 comprehenders are generally more likely to rely on heuristics-based processing strategies than L1 comprehenders, we would expect to observe *even slower* overall reading times with passive sentences in the L2 group in comparison to the L1 group.

Participants

Twenty-five L1 German, L2 speakers of English (age range = 18-47, mean age = 26, 19 female) and a control group of twenty-one L1 speakers of English (age range = 18-56, mean age = 27, 18 female) took part in this study. The L1 German participants' average age of exposure to English was eight years (range = 4-13) and all reported German as their L1 and dominant language, with the exception of two participants who reported German as their L1 and English as their dominant language. The L1 English participants all reported English as their L1 and dominant language. All participants reported normal or corrected-to-normal vision and no history of language or reading disorders were reported. Participants were recruited through Prolific (https://prolific.co/) and through contacts within the L1 community of interest. Participants recruited through Prolific were offered a small financial incentive following completion of the experiment.

English Proficiency

Since individual differences in proficiency may affect L2 learners' performance, following the main experiment, L1 German comprehenders completed the pen and paper section of the Oxford quick placement test (Oxford University Press, 2001), adapted to a computer-based format. Each correct answer carried one mark and participants' level of English proficiency was determined using the corresponding levels of the Common European Reference Framework for languages. The participants' levels of proficiency in English were categorised as lower intermediate (n=1), upper intermediate (n=3), lower advanced (n=12), and upper advanced (n=9).

Experimental Materials

The experimental materials consisted of 28 sentence sets, each with four sentences, one per condition, as per Table 1. To minimise experiment-specific task effects, 32 filler sentences were included in the experiment from an unrelated study investigating how readers re-access a previously mentioned noun when reading pronouns. The complete list of experimental and filler items are publicly available at the Open Science Framework (https://osf.io/rzmds/)

One feature of the study which was used as fillers for the current experiment required the inclusion of a context sentence preceding each target experimental sentence. This sentence was presented in its entirety on the screen, and participants were instructed to press a button to proceed to the target sentence (presented in a maze-task fashion) when they were ready. To prevent participant bias, simple declarative context sentences were also created for the present study. The sentences were designed so that they did not explicitly mention the referents in the experimental sentences and could reasonably introduce each sentence in the set. For example, as shown in Table 1, the context sentence *skill is essential in the hospitality industry* precedes each of the four experimental sentences.

For each target experimental sentence in the maze task, participants were given the first word, after which they had to choose between two words: the correct continuation in the sentence and a grammatically implausible foil word, as illustrated in Figure 1. Foil words were generated using the A-Maze procedure with the default Gulordava model (Boyce et al., 2020; Gulordava, Bojanowski, Grave, Linzen & Baroni, 2018), and hand-checked to ensure implausibility.

To measure offline processing, comprehension questions targeting thematic roles followed each sentence, as Good-Enough or shallow processing accounts predict that participants will frequently fail to correctly assign thematic roles following non-canonical passive sentences. Should L2 comprehenders be more inclined to rely on shallower processing heuristics, we would therefore expect to observe a smaller proportion of correct responses to comprehension questions following passive sentences in the L2 group than the L1 group.

Procedure

The experiment was implemented and hosted in Ibex Farm (Drummond, von der Malsburg, Erlewine & Vafaie, 2016). After consenting to the study and providing demographic information, participants entered a short session consisting of three practice trials to familiarise them with the experimental procedure. Data from the practice session was excluded from all analyses.

At the start of each trial, participants were presented with the context sentence and could take as long as required to read it before pressing the 'E' key to start the trial. Reading time data for the context sentences was not included in analyses. During the trial, each word of the sentence appeared alongside an implausible foil. Aside from the first word of the sentence which always appeared on the left of the screen, each subsequent word appeared in a randomised pattern to minimise the risk of artificially influencing participants' reading strategies.

After the last word of the sentence was revealed, the sentence disappeared, and participants were presented with the comprehension question to which they responded yes or no by using the 'E' or 'I' keys. The comprehension questions were set to timeout after 60,000ms, as any longer than this was presumed to indicate inattention. If at any point during the trial participants selected a foil word, the sentence would end, and the comprehension question would appear. Participants were encouraged to make their best guess based on how much of the sentence they had read. The practice session was followed by a randomised presentation of the 28 experimental sentences and 32 filler sentences. Participants were encouraged to take a break every 16 sentences to prevent fatigue or disengagement with the task. The presentation of the experimental sentences was randomised across trials. Including pre-test information, demographic forms, proficiency testing and post-test debriefing, the experiment took around 45 minutes to complete for the L2 group and around 30 minutes for the L1 group, who did not take part in the proficiency test.

Data Analysis

Prior to analysis, the complete reading time dataset consisted of 7056 tokens from twenty-one participants in the L1 English group, and 8400 tokens from twenty-five participants in the L2 English group. The question response dataset consisted of 529 tokens for analysis in the L1 group and 601 tokens in the L2 group. All data was analysed using RStudio (RStudio Team, 2021).

Both reading time & question response time data below 100ms and above 12,000ms were treated as outliers and excluded from analyses, under the assumption that faster than 100ms reflects erroneous button-presses, and slower than 12,000ms reflects inattention. Additionally, samples in which no reading time or reaction time (RT) were recorded were also excluded from analysis. Individual trials in which participants failed to reach the head of the second NP were also excluded from all further analysis, as it was imperative that participants saw at least the head of the second NP in the experimental sentence in order to answer the comprehension question. In line with Paolazzi et al. (2019, 2021), trials in which participants failed to answer the comprehension question correctly were also removed from further analysis. The rationale for this is that as a Good-Enough processing account assumes that greater accuracy is indicative of deeper, more complete sentence processing, then passives are most likely to exact a cost in these trials. Thus, excluding incorrect responses from the dataset allows shallow processing approaches the optimal chance to succeed. The filtered reading time data was then grouped by speaker to identify the mean reading times for each region of interest (ROI) before being further condensed by word, condition, and native language to identify the grand mean, standard deviation, and standard error. In total, the reading time data consisted of 6347 tokens for analysis in the L1 group and 7188 tokens in the L2 group. The filtered question response dataset was also grouped by speaker to identify the mean RT for each condition, before being condensed by condition and native language to identify the key summary statistics. In total, the question response dataset consisted of 522 tokens for analysis in the L1 group and 591 tokens in the L2 group.

Reading time and RT data were analysed using mixed effects regression modelling taking subjects and items as random effects grouping terms. The contrasts used were predicate (eventive= -1, stative= 1), syntax (active= 1, passive = -1), and L1 (English= -1, German= 1). The models were then backward-selected to find the maximal model supported by the data, thus allowing us to capture any regularities in inter-subject and item variability (Barr, Levy, Scheepers & Tily, 2013). Following Gelman and Hill (2006), *p*-values were generated by regarding the *t*-value as a *z*-score. The maximal model structure is given in (9), and the reduced model in (10).

(9) RT ~ syntax*predicate*L1 + (1+syntax*predicate*L1 | item) + (1+syntax*predicate|subj)

(10) RT ~ syntax*predicate*L1 + (1+predicate | subj) + (1+predicate+syntax:predicate | item)

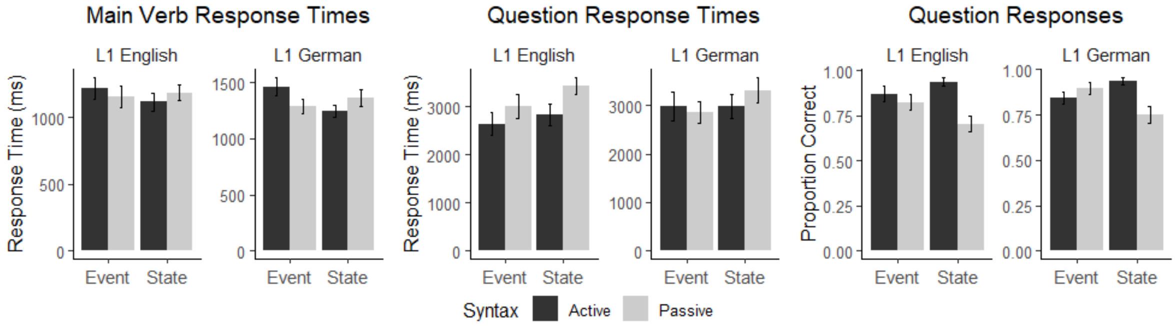
Comprehension accuracy data was analysed using a mixed effects logistic regression with syntax (active= 1, passive = -1), predicate type (eventive= -1, stative= 1), and native language (English= -1, German= 1) as fixed effects (including random slopes) and random intercepts on a by-subject and by-item basis. As above, the model was then backward-selected to find the maximal model supported by the data.

Results

Figure 2 summarises the key statistics for both group's reading times at the main verb, response times to comprehension questions, and proportion of correct responses in each condition. As seen in Figure 2, the L1 group read the main verb in eventive predicates more quickly in the passive condition than the active condition. However, in stative predicates, passives were read marginally slower than actives. The L2 group displayed a similar reading pattern: in eventive predicates, passives were read more quickly on average than actives . Similarly, for stative predicates, L2 speakers slowed down at the main verb in passives compared to actives.

Figure 2

By-subject means. Error bars represent standard error.



Turning to response times to comprehension questions, the L1 group was quicker to correctly respond following eventive predicates in the active condition than passives. Similarly, for stative predicates, L1 participants were also faster to correctly answer questions following actives than passives. A different pattern is observed in the L2 group, whereby participants were marginally quicker to respond correctly to questions following passives than actives in the eventive condition. However, concerning stative predicates, participants were faster to respond correctly to questions following actives than passives.

Considering the proportion of correct responses to comprehension questions, we again see slightly different patterns between groups. Following eventive predicates, the L1 group responded with marginally higher accuracy to questions following actives than passives. Concerning stative predicates, the L1 group responded with greater accuracy to questions following actives than passives. Following eventive predicates, however, the L2 group responded with greater accuracy to questions following passives than to those following actives. This pattern was reversed in the stative condition, with participants responding with greater accuracy to questions following actives than passives.

Table 2

Reading time data. Estimates for models taking syntax (active= 1, passive = −1), predicate (eventive= −1, stative= 1), L1 (English= −1, German= 1), and their interactions and fixed effects. Random slopes & intercepts were included for subjects and items. Bold typeface indicates significance.

	Main verb			**by**			**determiner**			**adjective 1**			**conjunction**			**adjective 2**			**noun**		
	$\hat{\beta}$	*t*-value	*p*-value	$\hat{\beta}$	*t*-value	*p*-value	$\hat{\beta}$	*t*-value	*p*-value	$\hat{\beta}$	*t*-value	*p*-value	$\hat{\beta}$	*t*-value	*p*-value	$\hat{\beta}$	*t*-value	*p*-value	$\hat{\beta}$	*t*-value	*p*-value
Syntax	5	0.27	.79		N/A		8	1.02	.31	−13	0.57	.57	**−35**	**2.81**	**<.01**	28	0.95	.34	−18	1.12	.26
Predicate	−28	1.18	.24	9.94	0.94	.35	−3	0.48	.63	5	0.27	.78	7	0.54	.59	6	0.25	.25	36	1.57	.12
L1	**89**	**4.92**	**<.01**	30.62	1	.32	**57**	**2.25**	**.02**	**89**	**1.98**	**.05**	51	1.55	.12	**143**	**3.35**	**<.01**	**123**	**3.17**	**<.01**
Syntax: Predicate	**−49**	**2.65**	**<.01**		N/A		−1	0.1	.92	11	0.5	.62	4	0.34	.73	4	0.17	.86	−13	0.62	.54
Syntax: L1	8	0.44	.66		N/A		8	0.92	.36	14	0.84	.4	24	*1.93*	*.05*	4	0.2	.84	2	0.14	.89
Predicate: L1	−8	0.46	.64	2.68	0.25	.8	**15**	**2.02**	**.04**	−25	1.47	.14	**27**	**2.17**	**.03**	8	0.51	.61	7	0.42	.67
Syntax: Predicate: L1	−15	0.82	.41		N/A		−4	0.58	.57	−1	0.05	.96	−14	1.12	.26	6	0.4	.69	−5	0.33	.74

As summarised in Table 2, we observed a significant interaction between syntax and predicate type in the reading time data at the main verb ($\hat{\beta}$ = −49, z = 2.65, p = <.01), indicating that across groups participants experienced selective difficulty with stative passives relative to eventive passives, and that eventive passives were read faster than eventive actives. The L2 comprehenders read both active and passive sentences overall more slowly than the L1 group, as demonstrated by a significant effect of native language at multiple ROIs, including the main verb ($\hat{\beta}$ = 89, z = 4.92, p = <.01), the determiner ($\hat{\beta}$ = 57, z = 2.25, p = .02), the first adjective ($\hat{\beta}$ = 89, z = 1.98, p = .05), the second adjective ($\hat{\beta}$ = 143, z = 3.35, p = <.01) and the noun ($\hat{\beta}$ = 123, z = 3.17, p = <.01). We also observed statistically significant interactions between native language and predicate type at multiple ROIs. At the determiner, the L1 English group read more quickly in the eventive conditions than the stative conditions ($\hat{\beta}$ = 15, z = 2.02, p = .04), whereas at the conjunction, the L2 English group read sentences in the eventive conditions more quickly than those in the stative conditions ($\hat{\beta}$ = 27, z = 2.17, p = .03). A main effect of syntax was also observed at the conjunction ($\hat{\beta}$ = −35, z = 2.81, p = <.01), such that passives were read more slowly than actives. Additionally, a trending interaction between syntax and language was observed, ($\hat{\beta}$ = 24, z = 1.93, p = .05) due to the L1 English group demonstrating a more uniform effect of syntax, taking longer to respond to passives than actives. This was not observed in the L2 English group. Crucially for the purposes of this study, we found no further evidence of an interaction between predicate type and syntax at any ROI other than the main verb.

Table 3

Question RT & accuracy data. Estimates for models taking syntax (active= 1, passive = −1), predicate (eventive= −1, stative= 1), L1 (English= −1, German= 1), and their interactions and fixed effects. Random slopes & intercepts were included for subjects and items. Bold typeface indicates significance.

	Question RT			**Accuracy**		
	$\hat{\beta}$	t-value	p-value	$\hat{\beta}$	t-value	p-value
Syntax	**−163**	**2.53**	**.01**	**0.59**	**5.44**	**<.01**
Predicate	**144**	**2.12**	**.03**	−0.19	1.13	.26
L1	18	0.13	.90	0.04	0.46	.65
Syntax:Predicate	−38	0.72	.47	**0.39**	**3.64**	**<.01**
Syntax:L1	**134**	**2.43**	**.01**	−0.13	1.32	.19
Predicate:L1	−37	0.79	.43	0.01	0.15	.88
Syntax:Predicate:L1	−30	0.65	.52	0.09	0.86	.39

Concerning the offline data, as summarised in Table 3, the question RTs revealed a statistically significant effect of syntax ($\hat{\beta} = -163$, $z = 2.53$, $p = .01$) due to participants taking longer to respond to questions following passives than actives. A statistically significant effect of predicate type was also observed ($\hat{\beta} = 144$, $z = 2.12$, $p = .03$) due to participants taking longer to respond to questions following stative predicates than eventives. Additionally, there was a statistically significant interaction between syntax and language ($\hat{\beta} = 134$, $z = 2.43$, $p = .01$), attributed to the L1 English group responding to questions following active sentences more quickly than those following passives. The comprehension accuracy data also revealed a statistically significant effect of syntax ($\hat{\beta} = 0.59$, $z = 5.44$, p = <.01) due to participants responding with greater accuracy to questions following actives than passives. A statistically significant interaction between syntax and predicate type was observed ($\hat{\beta} = 0.39$, $z = 3.64$, $p =$ <.01) due to the fact that syntax played a larger role for stative predicates than for eventive predicates.

In order to establish if there was statistically significant within-group variation as a function of syntax, predicate type, and English proficiency, a separate analysis was conducted on the L2 data alone, including centred proficiency as a continuous predictor in addition to syntax and predicate type. Reading times and comprehension question response times were analysed using mixed effects linear regression models. The contrasts used were syntax (passive −1, active 1), predicate type (eventive −1, stative 1) and centred proficiency. Centred proficiency was calculated by subtracting the mean proficiency score from each individual's proficiency score. Comprehension accuracy was analysed using mixed effects logistic regression, using the same contrasts and centred proficiency.

We failed to find evidence of an effect or interaction with proficiency in the reading time and comprehension question response time data. A statistically significant effect of proficiency was observed for comprehension accuracy ($\hat{\beta} = 0.09$, $z = 3.07$, $p = <.01$) led by participants with higher proficiency responding with greater accuracy to comprehension questions than lower proficiency participants, however this did not appear to interact with the main predictor variables of syntax and predicate type. For the complete proficiency data, see Appendix A.

Discussion

The aim of this study was to replicate the results from Paolazzi et al. (2019, 2021) using a different methodology, and to investigate whether these results could be extended to L2 speakers of English. The study contrasted syntax and predicate type, measuring online reading times in a maze task and offline effects through comprehension questions targeting thematic roles, measuring reaction times and accuracy. Our data adds to the findings of Paolazzi et al. (2019, 2021) showing that the L1-processing of passives is not inherently more complex than actives, but rather appears to be modulated by predicate semantics. We found complimentary results in our L2 group, suggesting that L2 comprehenders were no more inclined to rely on shallower processing mechanisms than L1 comprehenders, as suggested by the Shallow Structure Hypothesis.

Regarding the online data, the main verb was the only region to reveal a significant interaction between syntax and predicate type. Participants spent longer reading the main verb in the stative passive condition, however, we see no evidence suggesting lingering effects moving into the determiner region and through to the NP. This is to be expected given that the maze task means readers have to input information incrementally in order to choose the correct sentence continuation, and our data seems to suggest that passivization is modulated by event type. A main effect of language was found at almost every ROI in the online data, due to the L1 group reading overall more quickly than the L2 group. However, this finding alone is neither surprising nor unexpected given that L2 speakers will be generally less proficient than L1 speakers, and that proficiency is expected to influence processing cost and time. Additionally, it is worth considering that the L1 readers' overall faster reading times are likely to have reduced the amount of variation in the reading time data. For the purposes of this study, the important takeaway from the online data is that there was no evidence of an interaction between syntax and language, or syntax, language, and predicate type, suggesting that the way in which the two groups processed passive and active sentences in real time was not qualitatively different.

That we found no evidence of a difference between our L1 and L2 groups suggests that our L2 group did not demonstrate a positive transfer effect from their L1. Recall that in this case, we would have expected to see a larger comprehension penalty for passivised stative predicates in the L2 group in comparison to the L1 group. One possibility is that in moving from L1 German to L2 English, it is simpler to collapse categories (i.e. different auxiliaries) than it would be to learn new grammatical distinctions. Thus, relatively little transfer is observed because the transfer would be in the wrong direction. However, we recognise that our study was limited to L1 German speakers, therefore we remain cautious in interpreting these findings: future research would benefit from recruiting participants from typologically different language backgrounds in order to elucidate the role of L1 transfer. Specifically, gathering data on the processing patterns of those whose native languages do and do not routinely distinguish between passivised stative and eventive predicates would allow us a better understanding of how this distinction is acquired in L2 English.

Whilst no statistically significant interactions between syntax and language were observed in the online data, a trending interaction was observed at the conjunction which is worthy of discussion. Notably in this ROI, the L1 group read actives more quickly than passives, whereas the L2 group read eventive predicates more quickly than stative predicates, irrespective of voice. This differential language impact could potentially be considered evidence in support of shallow processing. However, moving into the next word in the sentence, the second adjective, the L2 group processed passives more quickly than actives. Assuming that the data at the conjunction is the result of shallow processing, the pattern at the second adjective contradicts this assumption. Alternatively, the different pattern between the L1 and L2 groups could be seen to suggest an effect of late processing measures in the L1 group only. Though it is uncertain why we would see this effect at the conjunction given that spillover effects are attenuated in the maze task and the conjunction is separated from the critical verb region by the by-phrase, determiner and first adjective. At present, the reason for the difference between the two groups at the conjunction remains unclear, though we suggest it is likely due to noise in a relatively small sample.

Concerning the offline data, the L1 group's data is consistent with the existing literature, in that passives elicited longer reaction times and lower overall accuracy, irrespective of predicate type. The L2 group, however, responded more quickly and with greater accuracy to passives than actives in the eventive conditions, with the opposite pattern observed for stative predicates. It does not appear that these results are consistent with a shallow processing account, as faster reaction times should be accompanied by a lower proportion of correct responses - indicative of shallow processing - and we should see the same pattern in passives, regardless of predicate type. It is possible that the difference observed in syntax was due to task-based effects. That is, as the task required constructing and reconstructing each and every sentence, it is possible that the L2 group found this task inherently more difficult than the L1 group. Given that stative passives require more repair than eventive passives, this could explain the differential predicate behaviour for the L2 group. However, that the L2 group was consistently able to do this, and with high overall accuracy, we take to suggest that shallow processing is not evidenced in the data.

The only measure in the offline data to reveal a statistically significant interaction between syntax and predicate type was accuracy, driven by the fact that participants responded more accurately to comprehension questions following actives than passives in the stative condition. In the absence of an interaction with L1, we have no evidence that language background interacted with these factors. Of note, both groups also took the longest to respond correctly to questions following passivised stative predicates, though this failed to reach statistical significance. Taken together, our results appear consistent with Paolazzi et al. (2019, 2021); suggesting that the ambiguity and coercion required to reach the adjectival interpretation of passivised stative predicates is more costly to online and offline processing resources in both L1 and L2 speakers of English.

That participants in both groups appeared to experience greater online and offline processing difficulties with passivised stative predicates, but not eventive predicates, is a potential challenge for Good-Enough and shallow processing accounts. Specifically concerning the Shallow Structure Hypothesis, should L2 speakers be more inclined to rely on a shallower processing strategy, we would not expect to see an effect or interaction with predicate semantics. Indeed, we would expect to see *much* lower overall accuracy to passives due to *consistent* failure to correctly assign thematic roles. However, the L2 group responded fastest, and with greater accuracy, to comprehension questions following passivised eventive predicates than those in the eventive active condition. Taken with the fact that L2 participants also read the main verb more quickly in passivised eventive predicates than statives, this suggests that not only are L2 speakers able to consistently build accurate representations of passive sentences during online processing, but they are able to recall and rebuild these structures in order to answer the offline comprehension questions. That we see both L1 and L2 English speakers reading the main verb more slowly in the stative passive condition and responding more slowly and with lower accuracy to questions following this construction seems to suggest that building a complex event structure from a stative predicate may be the source of difficulty. Additionally, it is worth noting that we see a relatively low overall proportion of incorrect responses in both groups, with the lowest accuracy to questions following passivised stative predicates: 70% accuracy in the L1 group and 75% accuracy in the L2 group. This effect is somewhat unexpected under a standard good-enough processing approach. Recall that it was this condition which caused participants to slow down while reading passive sentences. Under the assumption that slower reading times (at least in part) index the algorithmic processing, we would expect this condition to show correspondingly better comprehension, contrary to what is shown in the analyses above. That said, it would likely be possible to amend good-enough approaches to account for this asymmetry by, for example, more thoroughly articulating the kinds of heuristics being engaged by the parse, and the point(s) at which they are abandoned in favour of algorithmic processing.

Finally, no effects of proficiency were observed in the L2 group's reading time or response time data. A significant effect was observed in comprehension accuracy, such that less proficient speakers were less accurate in their responses. These findings, again, are unsurprising in light of the relatively small sample size and given that only four of the twenty-five German participants were categorised as below lower-advanced proficiency. Once again, we must acknowledge the limitations of the statistical power of the study and consider that we may lack sufficient data to observe an effect of proficiency. That is, our results are driven by the more highly proficient speakers of English, therefore we must question how more evenly-distributed groups would impact on the results. Specifically comparing lower proficiency speakers with higher proficiency speakers would have undeniable theoretical and practical applications, allowing us a deeper understanding of how L2 sentence processing develops with proficiency. Finally, in the future we aim to compare the current results with results from L1 speakers of languages typologically further away from English than German.

Conclusion

Our study adds to previous results showing that the L1 processing of passives is not inherently more complex than actives, using a different methodology and population. In line with the existing literature (Paolazzi et al., 2019; 2021), we show that the complexity of passivization is modulated by predicate semantics. Furthermore, we found no evidence of a difference between L1 and L2 English speakers, with both populations showing a nuanced effect of the interaction of event structure and passivization.

References

Abbot-Smith, K., & Behrens, H. (2006). How known constructions influence the acquisition of other constructions: The German passive and future constructions. *Cognitive Science*, *30*(6), 995–1026. https://doi.org/10.1207/s15516709cog0000_61

Bader, M., & Meng, M. (2018). The misinterpretation of noncanonical sentences revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(8), 1286–1311. https://doi.org/10.1037/xlm0000519

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Borer, H. & Wexler, K. (1987). The maturation of syntax. In Tomas Roeper & Edwin Williams (eds.), *Parameter setting*, (pp. 123–172). Dordrecht: Reidel.

Boyce, V., Futrell, R., & Levy, R. P. (2020). Maze Made Easy: Better and easier measurement of incremental processing difficulty. *Journal of Memory and Language*, *111*, 104082. https://doi.org/10.1016/j.jml.2019.104082

Carrithers, C. (1989). Syntactic complexity does not necessarily make sentences harder to understand. *Journal of Psycholinguistic Research*, *18*(1), 75–88. https://doi.org/10.1007/bf01069048

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic Roles Assigned along the Garden Path Linger. *Cognitive Psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Christianson, K., Luke, S. G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 538–544. https://doi.org/10.1037/a0018027

Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3–42. https://doi.org/10.1017/s0142716406060024

Clahsen, H., & Felser, C. (2006b). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*(1), 107–126. https://doi.org/10.1017/s0142716406060206

Clahsen, H., & Felser, C. (2006c). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*(12), 564–570. https://doi.org/10.1016/j.tics.2006.10.002

Clahsen, H., & Felser, C. (2018). Some notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition*, *40*(3), 693–706. https://doi.org/10.1017/s0272263117000250

Drummond, A., von der Malsburg, T., Erlewine, M. Y., & Vafaie, M. (2016). *Ibex farm*. https://github.com/addrummond/ibex

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, *24*(3), 453–489. https://doi.org/10.1017/s0142716403000237

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Ferreira, F., & Christianson, K. (2016). Is Now-or-Never language processing good enough? *Behavioral and Brain Sciences*, *39*. https://doi.org/10.1017/s0140525x15000771

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. https://doi.org/10.1016/s0010-0285(03)00005-7

Ferreira, F., & Patson, N. D. (2007). The "Good Enough" Approach to Language Comprehension. *Language and Linguistics Compass*, *1*(1-2), 71–83. https://doi.org/10.1111/j.1749-818x.2007.00007.x

Gehrke, B. & Grillo, N., (2007). Aspects on Passives. In: Blaho, Sylvia, Schoorlemmer, Eric and Vicente, Luis, (eds.) *Proceedings of ConSOLE XIV*. (pp. 121-141). Student Organisation of Linguistics in Europe.

Gehrke, B., & Grillo, N. (2009). How to become passive. *Explorations of Phase Theory: Features and Arguments*, 231–268.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/ hierarchical models.* Cambridge, MA: Cambridge University Press.

Grillo, N., Alexiadou, A., Gehrke, B., Hirsch, N., Paolazzi, C., & Santi, A. (2019). Processing unambiguous verbal passives in German. *Journal of Linguistics*, *55*(3), 523–562. https://doi.org/10.1017/s0022226718000300

Gulordava, K., Bojanowski, P., Grave, E., Linzen, T., & Baroni, M. (2018). Colorless green recurrent networks dream hierarchically. *North American Chapter of the Association for Computational Linguistics.* https://doi.org/10.18653/v1/n18-1108

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In L. Levin (Ed.), *Proceedings of the second meeting of the north American chapter of the association for computational linguistics* (pp. 159–166).

Jacob, G., & Felser, C. (2016). Reanalysis and semantic persistence in native and non-native garden-path recovery. *Quarterly Journal of Experimental Psychology*, *69*(5), 907–925. https://doi.org/10.1080/17470218.2014.984231

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, *69*(5), 1013–1040. https://doi.org/10.1080/17470218.2015.1053951

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Meng, M., & Bader, M. (2020). Does comprehension (sometimes) go wrong for noncanonical sentences? *Quarterly Journal of Experimental Psychology*, *74*(1), 1–28. https://doi.org/10.1177/1747021820947940

Osterhout, L., & Swinney, D. A. (1993). On the temporal course of gap-filling during comprehension of verbal passives. *Journal of Psycholinguistic Research*, *22*(2), 273–286. https://doi.org/10.1007/bf01067834

Oxford University Press, University of Cambridge, & Association of Language Testers in Europe. (2001). Quick placement test: Paper and pen test. Oxford: Oxford University Press.

Paolazzi, C. L., Grillo, N., Alexiadou, A., & Santi, A. (2019). Passives are not hard to interpret but hard to remember: evidence from online and offline studies. *Language, Cognition and Neuroscience*, *34*(8), 991–1015. https://doi.org/10.1080/23273798.2019.1602733

Paolazzi, C. L., Grillo, N., Cera, C., Karageorgou, F., Bullman, E., Chow, W. Y., & Santi, A. (2021). Eyetracking while reading passives: an event structure account of difficulty. *Language, Cognition and Neuroscience*, *37*(2), 135–153. https://doi.org/10.1080/23273798.2021.1946108

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, *25*(4), 501–528. https://doi.org/10.1017/s0272263103000214

RStudio Team (2021). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, *32*(2), 299–331. https://doi.org/10.1017/s0142716410000421

Street, J. A., & Dąbrowska, E. (2010). More individual differences in language attainment: How much do adult native speakers of English know about passives and quantifiers? *Lingua*, *120*(8), 2080–2094. https://doi.org/10.1016/j.lingua.2010.01.004

Traxler, M. J., Corina, D. P., Morford, J. P., Hafer, S., & Hoversten, L. J. (2013). Deaf readers' response to syntactic complexity: Evidence from self-paced reading. *Memory & Cognition*, *42*(1), 97–111. https://doi.org/10.3758/s13421-013-0346-1

Witzel, N., Witzel, J., & Forster, K. (2011). Comparisons of Online Reading Paradigms: Eye Tracking, Moving-Window, and Maze. *Journal of Psycholinguistic Research*, *41*(2), 105–128. https://doi.org/10.1007/s10936-011-9179-x

Appendix A: Proficiency Data

Table A1

Reading time data. Estimates for models taking syntax (active= 1, passive = −1), predicate (eventive= −1, stative= 1), centred proficiency, and their interactions and fixed effects. Random slopes & intercepts were included for subjects and items. Bold typeface indicates significance.

	Main verb			**by**			**determiner**			**adjective 1**			**conjunction**			**adjective 2**			**noun**		
	$\hat{\beta}$	*t*-value	*p*-value	$\hat{\beta}$	*t*-value	*p*-value	$\hat{\beta}$	*t*-value	*p*-value	$\hat{\beta}$	*t*-value	*p*-value	$\hat{\beta}$	*t*-value	*p*-value	$\hat{\beta}$	*t*-value	*p*-value	$\hat{\beta}$	*t*-value	*p*-value
Syntax	8	0.31	.76		N/A		17	1.29	.20	−4	0.12	.90	−10	0.55	.58	29	0.69	.49	−13	0.49	.62
Predicate	−36	0.96	.34	12	1.03	.31	11	0.97	.33	−17	0.61	.54	**35**	**2.02**	**.04**	24	0.70	.49	42	1.29	.20
Proficiency	−6	1.24	.22	6	0.55	.58	8	0.88	.38	−2	0.15	.88	3	0.27	.79	−9	0.63	.53	5	0.34	.73
Syntax: Predicate	**−60**	**2.33**	**.02**		N/A		−5	0.43	.67	12	0.43	.67	−12	0.68	.49	−1	0.02	.99	−19	0.52	.60
Syntax: Proficiency	−2	0.48	.63		N/A		−3	1.25	.21	6	1.14	.26	−2	0.67	.50	2	0.27	.79	−9	1.73	.08
Predicate: Proficiency	−1	0.22	.82	1	0.25	.81	1	0.35	.72	7	1.20	.23	−5	1.27	.20	−5	0.96	.34	5	0.90	.37
Syntax: Predicate: Proficiency	6	1.25	.21		N/A		−1	0.25	.80	−1	0.17	.86	4	1.08	.28	7	1.37	.17	3	0.65	.51

Table A2

Question RT & accuracy data. Estimates for models taking syntax (active= 1, passive = −1), predicate (eventive= −1, stative= 1), centred proficiency, and their interactions and fixed effects. Random slopes & intercepts were included for subjects and items. Bold typeface indicates significance.

	Question RT			Accuracy		
	$\hat{\beta}$	*t*-value	*p*-value	$\hat{\beta}$	*t*-value	*p*-value
Syntax	−8.85	0.11	.91	**0.41**	**2.62**	**.01**
Predicate	88.59	1.27	.20	−0.11	0.56	.58
Proficiency	24.71	0.55	.58	**0.09**	**3.07**	**<.01**
Syntax:Predicate	−108.9	1.64	.10	**0.48**	**3.14**	**<.01**
Syntax:Proficiency	1.54	0.1	.92	0.02	0.84	.40
Predicate:Proficiency	5.79	0.42	.68	0.01	0.51	.61
Syntax:Predicate:Proficiency	17.38	1.28	.20	−0.05	1.85	.06