Syntactic structure frequencies do not predict observed approaches to syntactic repair of difficult garden-path sentences

Xulián Romano
Department of Linguistics
Department of Cognitive Science
Cornell University
xtr2@cornell.edu

Ethan Waxman
Department of Computer Science
Department of Applied
and Engineering Physics
Cornell University
ejw99@cornell.edu

Abstract

The study of garden-path sentence processing in humans is a window into the cognitive mechanisms of syntactic processing. Although previous work has shown that humans often fail to parse difficult garden-path sentences, little is known about how humans extract meaning from consequent incoherent parses. The current study investigates 1) whether humans add new words to incoherent parses of garden-path sentences to "fix" them, 2) which fix strategies humans prefer, and 3) what factors drive fix-strategy preferences. We use a forced choice task combined with a corpus frequency analysis to test whether preference for a fix strategy correlates with the frequency of the syntactic structures that result from that strategy. We conclude that neither the frequency of the syntactic structures that result from fix strategies *nor* the structural similarity between fixed garden-path sentences and their initial incoherent parses predict our results. Instead, we propose an explanation motivated by the complexity of thematic relations in fixed sentence types.

1 Introduction

1.1 Garden-path effects

Garden-path effects are increases in sentence reading time that occur when a reader begins to parse a sentence in a way that ultimately produces an ungrammatical result. Sentences that have locally ambiguous syntactic structures often elicit garden-path effects. The sentence in (1a) exhibits one of these localized syntactic ambiguities.

(1a) The dog walked by the bridge collapsed.

Readers of (1a) might initially interpret *walked* as the main verb of the sentence; however, once they reach *collapsed*, this interpretation is no longer viable because it ceases to be grammatical. The ambiguity that causes readers to misanalyze (1a) comes from the fact that 1) the past participle and past tense of the verb walk are homographs as well as homophones, and 2) relative pronouns (e.g., *who*) and copulas (e.g., *was*) can be omitted from passive object-extracted relative clauses to form reduced passive object-extracted relative clauses.

(1b) The dog (who was) walked by the bridge collapsed.

The sentence in (1a) contains a reduced passive object-extracted relative clause; the sentence in (1b) is the intended, grammatical interpretation of (1a). In this paper we refer to sentences that elicit garden-path effects, such as (1a), as garden-path sentences and the human syntactic processing system as the (human) parser.

1.2 Models of syntactic reanalysis

The cognitive mechanism by which the parser repairs syntactically anomalous sentences, such as garden-path sentences, is under investigation in the current study. Previous work on syntactic repair mechanisms focuses on syntactic reanalysis—the process by which the parser might *reanalyze* potentially ambiguous or ungrammatical sentences to find mistakes in their initial analyses (Bader, 1998; Gorrell, 1995; Pritchett, 1988; Sturt, 1997; Sturt and Crocker, 1996). Theoretical models of syntactic reanalysis differ in terms of how committed they are to preserving syntactic structural representations from these initial (mis)analyses. Some reanalysis models follow a principle of minimal structural revisions to initial parses; others apply human memory-inspired constraints on the kinds of structural modifications and alternative parses that the reanalysis mechanism can make (Sturt, 1997).

Another point of contention is the strategy by which reanalysis models search sentences for misanalyzed constituents and repair their structural representations. Three strategies have been proposed in the literature: backward reanalysis, forward reanalysis, and selective reanalysis. In backward reanalysis, the parser searches for the point of misanalysis by reading backward through the sentence and considering alternative syntactic structures for each phrasal constituent; in forward reanalysis, the parser searches for the point of misanalysis by rereading the sentence from the beginning while maintaining its initial incoherent parse for comparison with the corrected one; and in selective reanalysis, the parser searches for the point of misanalysis by selecting a region of the sentence to reanalyze based on information about the initial parse (Sturt, 1997).

1.3 Reanalysis and the brain

Functional magnetic resonance imaging (fMRI) has revealed that non-garden-path sentences—including garden-path sentences that have been disambiguated by pauses or prosodic cues—elicit more spread-out activity in the brain during aural language processing than do garden-path sentences (den Ouden et al., 2015). Aural processing of garden-path sentences is associated with brain activity that is mostly localized to the inferior frontal gyrus—a region that is subdivided into distinct sections that respond to syntactic-semantic mismatch detection, structural reanalysis, and structural complexity (den Ouden et al., 2015). Since syntactic reanalysis of garden-path sentences does not seem to recruit regions of the brain that handle non-language-specific tasks, it is unlikely that the reanalysis mechanism makes much use of non-linguistic information.

Electroencephalography (EEG) shows that garden-path processing triggers the event-related potential known as the P600 (Gouvea et al., 2010; Osterhout et al., 1994; Coulson et al., 1998). Magnetic encephalography (MEG) suggests that the P600 is generated in the posterior temporal lobe (Service et al., 2007)—a region associated with verb argument structural processing (den Ouden et al., 2015; Assadollahi and Rockstroh, 2008). The P600 is also associated with processing of ungrammatical sentences (Gouvea et al., 2010; Kaan and Swaab, 2003), complex but non-garden-path sentences (Gouvea et al., 2010; beim Graben et al., 2008; Kaan and Swaab, 2003), and semantically nonsensical but grammatical non-garden-path sentences (Kim and Osterhout, 2005). Since the P600 coincides with reanalysis, the range of syntactic anomalies that trigger the P600 suggests that the reanalysis mechanism is not specialized to handle a single type of syntactic anomaly.

1.4 Current study

Reanalysis cannot be understood fully without consideration for the situations in which it fails to build or preserve coherent structural representations. For example, readers sometimes retain the initial misanalysis of a garden-path sentence, even after correct reanalysis (Christianson et al., 2001). In other cases, the parser even fails to build coherent structural representations for sufficiently difficult garden-path sentences (Ceháková and Chromý, 2023).

With this consideration in mind, we adopt the position that reanalysis is a multi-stage process. In our framework, the parser first performs what we call analytic repair. This stage is inspired by traditional notions of reanalysis, especially the representation-preserving theoretical models described by Sturt (1997). If analytic repair fails to build a coherent structural representation for a sentence s, then the parser moves on to the next stage: synthetic repair. We define synthetic repair to be the insertion of novel linguistic content—often in the form of disambiguating words—into the structural representation of s to make it cohere. The current study aims to investigate whether synthetic repair indeed occurs in cases of analytic repair failure on difficult garden-path sentences, and, if so, which kinds of synthetic repair operations humans perform.

One possibility is that if the parser fails to build a coherent structural representation for a garden-path sentence g, it assumes that a word is missing from g and swaps the structural representation of g out for the most frequent coherent syntactic structure that includes all nonsilent terminal nodes in g. We will refer to this hypothesis as the structure frequency hypothesis. To test the structure frequency hypothesis we use a forced-choice task in which participants choose between five "fix" strategies for difficult garden-path sentences and compare the frequency and preference trends of each strategy to the frequency of the syntactic structures that underlie each strategy in an English language corpus.

2 Experiment 1

2.1 Methods

2.1.1 Materials

Two hundred target items and 80 fillers were created. Target items consisted of a sentence stimulus similar to (2a) and 2 of 5 possible interpretations of that sentence.

(2a) The horse raced past the barn fell.

One possible interpretation was the unreduced version of the stimulus in (2b), and the others, which we call fix strategies, were sourced from an informal preliminary survey in which 16 native speakers of English were asked to give their interpretation of (2a). The survey responses in (2c-f) inspired the fix strategies in our survey.

(2b) The horse **that was** raced past the barn fell.
Intended parse (IP)
(2c) The horse **that** raced past the barn fell.
Subject-modifier relative clause (SR) fix
(2d) The horse raced past the barn **and** fell.
Conjoined VP (CV) fix
(2e) The horse raced past**, and** the barn fell.
Conjoined sentence (CS) fix
(2f) The horse raced past the barn **that** fell.
Object-modifier relative clause (OR) fix

Filler stimuli consisted of 20 sentences similar to (3a) that exhibited prepositional phrase (PP) attachment ambiguity and 20 sentences similar to (4a) in which the main verb phrase (VP) had a reduced sentential complement.

(3a) The tourist saw the man with the binoculars.

Unlike the target stimuli, the type of sentence in (3a) exhibits a global ambiguity (i.e., who has the binoculars), which cannot be resolved without contextual cues.

(4a) The sailor knew the captain was ill.

Unlike the target stimuli, the type of sentence in (4a) is unambiguous, but it has a clearer, unreduced form (i.e., the sailor knew that the captain was ill).

To parallel the target items, fillers also displayed 2 possible interpretations of the sentence stimulus: a genuine interpretation similar to (3b) and (4b) and a dummy interpretation similar to (3c) and (4c). The dummy interpretations were included to discourage participants from overthinking the task and designed to mimic patterns in the target fix strategies.

(3b) The tourist saw the man **through** the binoculars.
(3c) The tourist saw the man **and** the binoculars.

(4b) The sailor knew **that** the captain was ill.
(4c) The sailor knew**, and** the captain was ill.

2.1.2 Design and procedure

Participants were asked to read sentences and choose the best of 2 possible interpretations (fix strategies) for each. The experiment was programmed in PCIbex. The PCIbex code recorded the interpretation option that participants selected for each item as well as their response time. Target items each displayed 1 of 20 possible permutations of fix strategy options. Each participant saw 20 of the 200 possible target item configurations. To eliminate a possible option-order confound, every participant was exposed to the 20 permutations of fix strategy options across their 20 target items. The order of the items was randomized for each participant.

2.1.3 Participants

Thirty-two native speakers of American English were recruited; each gave informed consent to participate in the study. Three responses in 3 participants' results were excluded due to technical issues during data collection that affected those responses.

2.2 Predictions

The structure frequency hypothesis predicts that participants would prefer the fix strategy with the most frequently occurring syntactic structure. According to Experiment 2, this would mean a preference for the CS fix. The informal fix-strategy survey used to design the materials in this experiment predicts that participants would prefer the CV and SR fixes; 7 survey participants generated CV fixes, 5 generated SR fixes, 2 generated OR fixes, and 2 generated CS fixes.

2.3 Results

The fix strategy with the highest frequency across all trials (overall frequency in Figure 1) was SR (23.9%, $p < 0.01$). The fix strategy with the lowest overall frequency was OR (16.3%, $p < 0.01$). Significance tests of proportion did not find the differences in overall frequency between the remaining fix strategies to be significant. CV had the second highest overall frequency (21.4%, $p < 0.20$) followed by CS (20.3%, $p < 0.44$). The intended parse had the second lowest overall frequency (18.2%, $p < 0.13$).

Head-to-head pairings of fix strategies (Figure 2) show a significant preference across participants for SR over the IP (70%, $p < 0.002$) and OR (62%, $p < 0.02$). SR is also preferred over CS

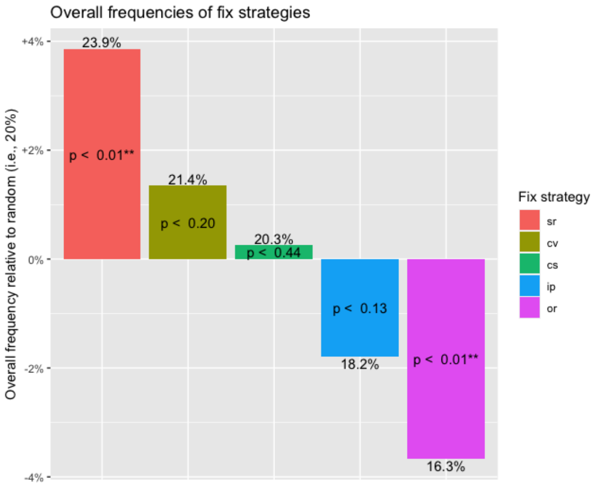

Figure 1: Overall frequencies of fix strategies compared relative to a baseline of chance (20%)

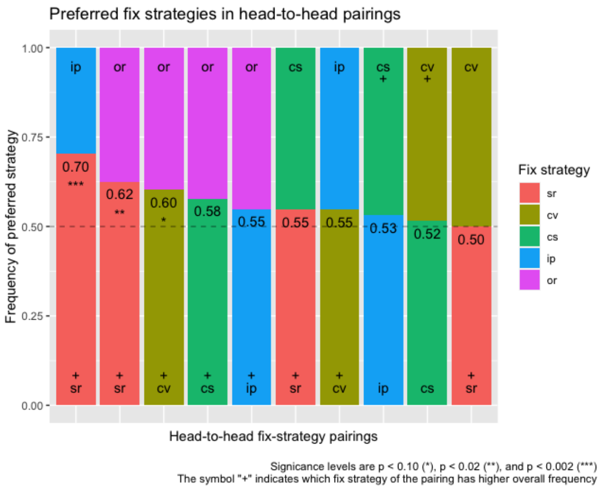

Figure 2: Selection rates of fix strategies when paired on experimental items

(55%). There is no preference in the SR-CV pairing—exactly 50% of participants chose each option when they encountered this pairing. In no pairings is OR preferred. Pairings with OR exhibit the highest rates of preference for CV (60%, $p < 0.10$), CS (58%), and the IP (55%).

The SR-OR pairing was sensitive to the order in which participants read the interpretation options (see Figure 3). If SR was read first, it was preferred over OR 53% of the time, but if SR was read second, its rate of preference went up to 72%; this is a near significant effect ($p < 0.06$). No other order effects were significant, but the second largest was observed in the CS-OR pairing. The order effect in CS-OR is similar to the order effect SR-OR, but the rate of CS preference increases from 50% to 66% ($p < 0.10$).

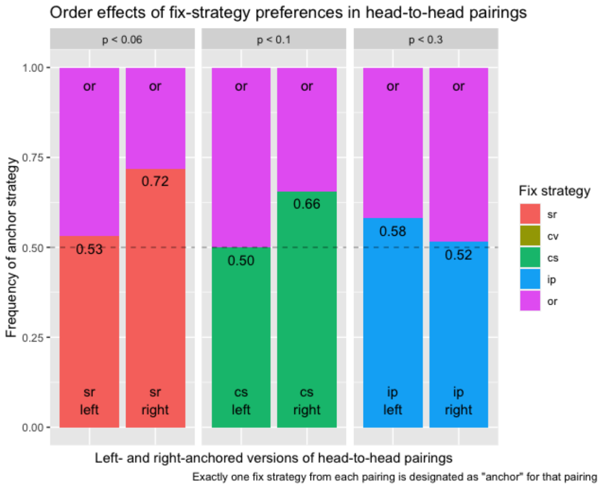

Figure 3: Order effects of fix strategies as options on experimental items

2.4 Discussion

In terms of overall frequency (Figure 1) and preference in head-to-head pairings (Figure 2), the SR fix is preferred over the IP (as well as every other fix strategy). This preference hierarchy is predicted by the results of the pre-experimental survey in which SR was among the most frequent fix strategies.

Since participants never exhibited a significant preference for the IP over other fix strategies, it is unlikely that they recognized this fix's relationship with the sentence stimuli. This suggests that participants failed to build coherent structural representations for the stimuli even when the IP was an option. One explanation for this result is that the IP often requires participants to make bizarre verb argument assignments. Consider for example the following sentences:

(5a) The man scooped the ice cream slipped.
(5b) The man (who was) scooped the ice cream slipped.
(5c) The man who scooped up the ice cream slipped.

In (5b), the verb *scoop* takes two noun phrase (NP) arguments, where *man* receives the θ-role of goal; usually the object of *scoop* is a PP goal, as demonstrated in (5c). Even if participants recognized (5a) as a reduced passive object-extracted relative clause, they may have disqualified that parse because of the uncharacteristic θ-relationship between *man* and *scoop*.

Of particular note is the fact that syntactic structural similarity between fix strategies does not correlate with fix-strategy preferences. For example, the SR fix and the OR fix result in fixed sentences with similar syntactic structures, but there is an extremely strong preference for SR over most other fix strategies and an equally strong aversion to OR. This suggests that representation preservation may not be as important in synthetic repair as Sturt (1997) argues it to be in analytic repair. Additionally, no fix-strategy preference is observed in the SR-CV pairing—in which, both options are identical in meaning. This might suggest that semantic factors have a significant influence on fix strategy preferences [1].

3 Experiment 2

3.1 Methods

3.1.1 Materials

To test the structure frequency hypothesis, an investigation into the structure frequencies of the fix strategies was conducted. The neural language model used by van Schijndel and Linzen (2021)[2] was used to approximate the frequency of the syntactic structures of each fix strategy from Experiment 1. This model consists of 2 layers of long short-term memory (LSTM) recurrent units—200 units per layer—and was trained on an 80 million word subset of English Wikipedia with an average sentence length of 27 words. In keeping with van Schijndel and Linzen (2021), we will refer to this model as WikiRNN. Previous work has shown that WikiRNN is "sensitive to subject-verb agreement across intervening nouns (Gulordava et al., 2018), filler-gap dependencies (Wilcox et al., 2018), and constructions with temporary syntactic ambiguities (Frank and Hoeks, 2019; Futrell et al., 2019; van Schijndel and Linzen, 2018), among other syntactic phenomena" (van Schijndel and Linzen, 2021).

Forty test sentences—20 with a single high-frequency syntactic structure and 20 with a single low-frequency syntactic structure—were created to test whether the model would provide reliable measures of frequency for syntactic structures of the fix strategies. The high-frequency test structure—a sentence with no NP modifiers and a transitive main verb—can be seen in (6); the low-frequency test structure—a case of ineffective center-embedding—can be seen in (7).

(6) The nurse administered the vaccine.

(7) The man the cat the dog bit chased ran.

3.1.2 Procedure

Language model surprisal—negative log probability of a word conditioned on the preceding words—was extracted from the model for each word w in the test sentences as well as the fixed sentences (grouped by fix strategy) and garden-path sentences from Experiment 1.

$$\text{surp}(w_i) = -\log \mathbb{P}(w_i | w_{i-1}, w_{i-2}, \ldots, w_{i-n})$$

Sentence surprisal was defined as the mean surprisal of the words in a sentence s.

$$\text{sentsurp}(s) := \mu\big(\text{surp}(w_i) \mid w_i \in s\big)$$

Syntactic structure surprisal was defined as the mean surprisal of the sentences within each sentence group g.

$$\text{structsurp}(g) := \mu\big(\text{sentsurp}(s_i) \mid s_i \in g\big)$$

In keeping with the relationship between probability and surprisal, the following equation was used to calculate syntactic structure probability (frequency) for each sentence group:

$$\mathbb{P}(g) = \exp\big(-\text{structsurp}(g)\big)$$

The following equation was used to calculate the frequencies of selected syntactic structures relative to one another:

$$f(s) = \frac{\mathbb{P}(g)}{\sum \mathbb{P}(g_i)}$$

3.2 Predictions

Previous work on English syntactic structure frequency predicts that the syntactic structures of SR and OR would be less represented in most language model training corpora (Roland et al., 2007). If our method tracks syntactic structure frequency effectively, it should assign low frequencies the syntactic structures of SR and OR.

3.3 Results

The model-predicted relative frequencies of the high- and low-frequency test structures were 15.8% and 1.1% respectively. The model-predicted frequency of the garden-path syntactic structure from Experiment 1 was 8.6%. Among the fix strategies, CS had the highest model-predicted frequency (23.9%) followed by CV (17.1%) and IP (13.8%). SR (9.3%) and OR (1.1%) had the lowest model-predicted frequencies of the fix strategies.

[1] See 4.1

[2] The code for this model can be accessed at https://github.com/vansky/neural-complexity

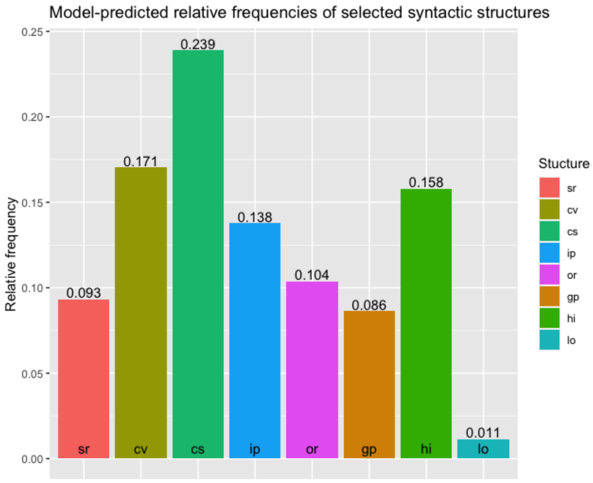

Figure 4: Garden-path sentences from Experiment 1 are labeled as GP; high- and low-frequency test sentences are labeled as HI and LO respectively.

3.4 Discussion

As a sanity check, the high- and low-frequency test structures have an appropriately high magnitude of difference in terms of the model-based assessment of their syntactic structure frequencies. The high-frequency test structure is similar in model-based frequency to the more frequent fix strategy structures, and the low-frequency test structure is extremely infrequent relative to all other syntactic structures in Experiment 2. Many of the model-based assessments of syntactic structure frequencies are supported by the literature on syntactic structure frequencies in English language corpora. By way of our frequency calculation method, WikiRNN produces similar results to Roland et al. (2007), who found that subject-extracted relative clauses (seen in SR and OR) occur with very low frequency. Our method also produces results that align with Roland et al. (2007) on the frequency of reduced passive object-extracted relative clauses (seen in GP), namely that they occur with higher frequency than non-passive subject-extracted relative clauses. The model also matches intuitionistic predictions of the frequency of CS.

In regard to the results from Experiment 1, the fix-strategy preferences of participants do not correlate with the model-based assessments of their respective syntactic structure frequencies. SR is the preferred fix strategy in Experiment 1, but the syntactic structure associated with SR is predicted by our method to be the least frequent structure in the training corpus. Similarly, the OR structure, which corresponds to the least preferred fix strategy in Experiment 1, is predicted to occur at a similar (slightly higher) rate to the SR structure in the training corpus. These results do not support the idea that fixes depend on the frequency of syntactic structures in a language. In 4.1, we propose an alternative explanation for the fix strategy preferences observed in Experiment 1.

4 General discussion

4.1 Future areas of interest

In 2.4, we suggested that factors from the semantic level of representation might influence fix strategy preferences. Here we develop this idea further and describe what might be a semantically conditioned pattern in the results of Experiment 1. These results show an inverse correlation between the number of entities performing actions in a fixed sentence s and preference for the fix strategy that produced s. In other words, a fix strategy is more likely to be applied the fewer distinct NPs receive agent-like semantic roles (e.g., agent or experiencer) as result of its application.

Single NP agent: One entity performs a series of actions

(SR) The horse that raced past the barn fell.
(CV) The horse raced past the barn and fell.

Multiple NP agents: Two entities perform distinct actions

(CS) The horse raced past, and the barn fell.
(IP) The horse who was raced past the barn fell.
The presupposed jockey is the second entity.
(OR) The horse raced past the barn that fell.

This suggests that syntactic fix strategies are governed by a principle of semantic role-assignment parsimony. Another reason to explore semantic effects of this kind on synthetic repair is that the P600 event-related potential, which has been linked with reanalysis by way of its association with garden-path processing, is generated in the area of the brain that handles verb argument structure (den Ouden et al., 2015; Assadollahi and Rockstroh, 2008)—a domain of the syntax-semantics interface. Perhaps upon the parser's failure to reanalyze a syntactically anomalous sentence, it turns the task of sentence interpretation over to other modules, which employ semantically motivated syntactic fix strategies.

In this study, we assumed a framework for syntactic reanalysis that stipulated multiple mechanisms, namely analytic and synthetic repair. This assumption was not motivated by literature on multi-stage frameworks of syntactic reanalysis but rather the wish to separate fix strategies from traditional notions of syntactic reanalysis without conclusively divorcing the two. An alternative framework could be one in which fix strategies are not considered under the umbrella of reanalysis. Such a framework might provide a more adequate explanation for (what appears to be) the semantically conditioned nature of fix-strategy preferences. Further investigation is required to differentiate the mechanisms that reanalyze and rearrange syntax trees from those that perform insertion or deletion operations.

4.2 Conclusion

(8a) The horse raced past the barn fell.

(8b) The horse **that** raced past the barn fell. *Subject-modifier relative clause (SR) fix*

We investigate the mechanisms that underlie syntactic reanlysis, specifically those that come into play when the parser fails to find a coherent structural representation for difficult garden-path sentences modeled after (8a). We show that humans prefer syntactic fixes to these sentences that result in the syntactic structure seen in (8b), which we call the subject-modifier relative clause fix. We also show that nueral language-model surprisal—averaged over a set of sentences with the same syntactic structure—can reflect the frequency of that syntactic structure in the model's training corpus. Crucially, we combine these findings to demonstrate that the strategies humans use to fix incoherent sentence parses are not meaningfully related to the syntactic structure frequency of the consequent fixed parses. To conclude, we propose a possible semantic explanation for syntactic fix strategy preferences and a direction of further computational psycho- and neurolinguistc research into syntactic reanalysis.

Limitations

In our experiments, we encountered two notable limitations. First is that we did not gather enough human data to see whether some of the fix-strategy preferences measured in Experiment 1 were statistically significant. To investigate the frequency of fix strategies, we recruited 32 participants; 3 responses in 3 participants' results were excluded. While the most important results were statistically significant, it is possible that we missed some findings of interest due to a lack of data. Replicating Experiment 1 with more participants may serve to strengthen our findings.

The second limitation was that we were only able to use WikiRNN to generate model-based assessments of syntactic structure frequencies. WikiRNN was trained on a single, written language source that may not reflect the empirical frequency distribution of syntactic structures in English. An extended analysis could compare the results of WikiRNN with those of the SoapRNN from van Schijndel and Linzen (2021), which was trained on dialogue from soap operas, as well as various other models with distinct architectures and training corpora.

Ethics Statement

We believe that we have upheld the ACL Ethics Policy (outlined `https://www.acm.org/code-of-ethics`[here]) in this study. We have striven to do honorable, ethical work. The language model architecture and corpus we used in Experiment 2 were free, accessible, and open source. Human participants in Experiment 1 each gave informed consent prior to the study; the study came at no mental or physical cost to the participants, who were free to terminate their participation at any point during the experiment. The work we present is our own unless otherwise cited.

References

Ramin Assadollahi and Brigitte S Rockstroh. 2008. Representation of the verb's argument-structure in the human brain. *BMC Neurosci.*, 9(1):69.

Markus Bader. 1998. *Prosodic Influences on Reading Syntactically Ambiguous Sentences*, pages 1–46. Springer Netherlands, Dordrecht.

Peter beim Graben, Sabrina Gerth, and Shravan Vasishth. 2008. Towards dynamical system models of language-related brain potentials. *Cognitive neurodynamics*, 2:229–55.

Markéta Ceháková and Jan Chromý. 2023. Garden-path sentences and the diversity of their (mis)representations. *PLOS ONE*, 18(7):1–38.

Kiel Christianson, Andrew Hollingworth, John Halliwell, and Fernanda Ferreira. 2001. Thematic roles assigned along the garden path linger. *Cognitive psychology*, 42:368–407.

Seana Coulson, Jonathan King, and Marta Kutas. 1998. Expect the unexpected: Event-related brain response to morphosyntactic violations. *Lang. Cogn. Process*, 13:21–58.

Dirk-Bart den Ouden, Michael Dickey, Catherine Anderson, and Kiel Christianson. 2015. Neural correlates of early-closure garden-path processing: Effects of prosody and plausibility. *Quarterly journal of experimental psychology (2006)*, 8:1–58.

Stefan Frank and John Hoeks. 2019. The interaction between structure and meaning in sentence comprehension: Recurrent neural networks and reading times. In *Proceedings of the 41st annual meeting of the Cognitive Science Society (CogSci)*, pages 337–343, Austin, TX. Cognitive Science Society.

Richard Futrell, Ethan Wilcox, Takashi Morita, Peng Qian, Miguel Ballesteros, and Roger Levy. 2019. Neural language models as psycholinguistic subjects: Representations of syntactic state. In *Proceedings of the 2019 conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies, Vol. 1 (long and short papers)*, pages 32–42, Minneapolis, MN. Association for Computational Linguistics.

Paul Gorrell. 1995. *Syntax and Parsing*. Cambridge Studies in Linguistics. Cambridge University Press.

Ana Gouvea, Colin Phillips, Nina Kazanina, and David Poeppel. 2010. The linguistic processes underlying the p600. *LANGUAGE AND COGNITIVE PROCESSES*, 25:149–188.

Kristina Gulordava, Piotr Bojanowski, Edouard Grave, Tal Linzen, and Marco Baroni. 2018. Colorless green recurrent networks dream hierarchically. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1195–1205, New Orleans, Louisiana. Association for Computational Linguistics.

Edith Kaan and Tamara Swaab. 2003. Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of cognitive neuroscience*, 15:98–110.

Albert Kim and Lee Osterhout. 2005. The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52(2):205–225.

Lee Osterhout, Phillip J. Holcomb, and David Swinney. 1994. Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of experimental psychology. Learning, memory, and cognition*, 20 4:786–803.

Bradley L. Pritchett. 1988. Garden path phenomena and the grammatical basis of language processing. *Language*, 64(3):539–576.

Douglas Roland, Frederic Dick, and Jeffrey L. Elman. 2007. Frequency of basic english grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57(3):348–379.

Elisabet Service, Päivi Helenius, Sini Maury, and Riitta Salmelin. 2007. Localization of syntactic and semantic brain responses using magnetoencephalography. *J. Cogn. Neurosci.*, 19(7):1193–1205.

Patrick Sturt. 1997. *Syntactic Reanalysis in Human Language Processing*. Ph.D. thesis, University of Edinburgh.

Patrick Sturt and Matthew W. Crocker. 1996. Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes*, 11(5):449–494.

Marten van Schijndel and Tal Linzen. 2018. Modeling garden path effects without explicit hierarchical syntax. In *Proceedings of the 40th annual meeting of the Cognitive Science Society (CogSci)*, pages 2603–2608, Austin, TX. Cognitive Science Society.

Marten van Schijndel and Tal Linzen. 2021. Single-stage prediction models do not explain the magnitude of syntactic disambiguation difficulty. *Cognitive Science*, 45.

Ethan Wilcox, Roger Levy, Takashi Morita, and Richard Futrell. 2018. What do RNN language models learn about filler–gap dependencies? In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 211–221, Brussels, Belgium. Association for Computational Linguistics.

A Figure appendix

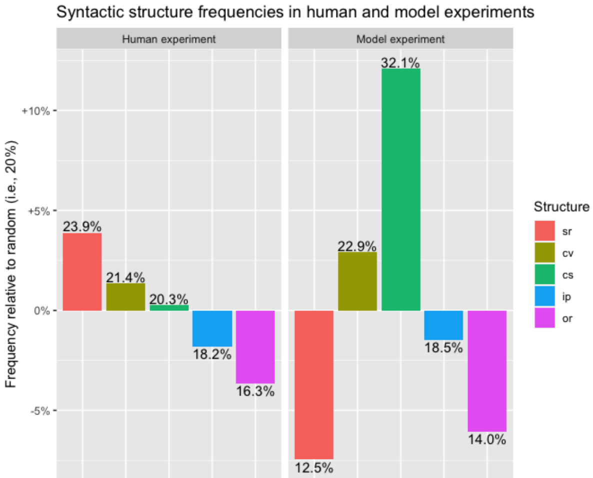

Figure 5: Distribution of fix-strategy syntactic structures in human and model experiments

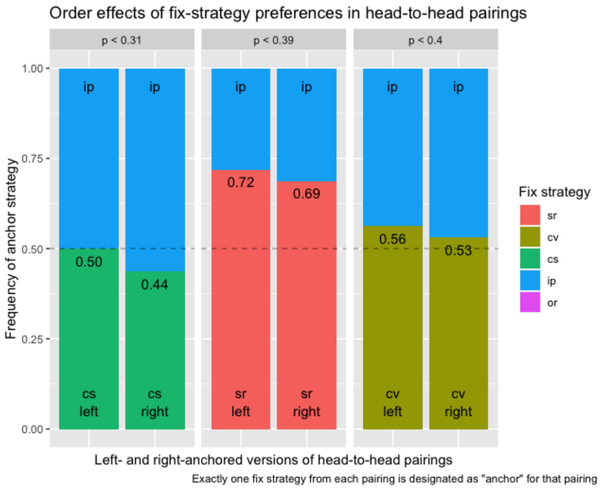

Figure 6: Order effects of fix strategies as options on experimental items

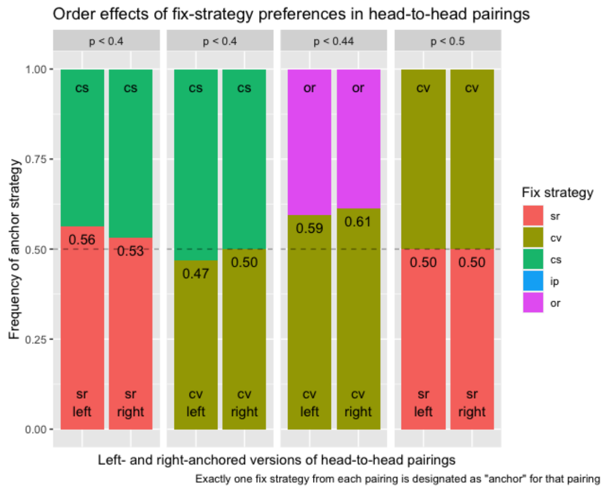

Figure 7: Order effects of fix strategies as options on experimental items