Review in Left Corner Parsing in Sentence Processing

Shisen Yue

Department of Linguistics, Shanghai Jiao Tong University

Abstract

How written words are encoded and decoded by the human brain? To pursue an answer to this question, we review on research in sentence processing and left corner parsing theories. We particularize the models and results in five studies that propose a left corner parser and conduct simulations with hand-crafted sentences or data from large-scale corpora. The high accuracy in simulations demonstrates both the theoretical and neurological significance of left corner parsing. Moreover, we discuss the inaccurate predictions of left corner parsing as a single predictor and the necessity of combining it with expectation-based theories.

Review in Left Corner Parsing in Sentence Processing

Introduction

The question of how sentences are processed in the human brain has received broad attention from researchers dedicated to both psycholinguistics and neurolinguistics. With the progress in computational modeling, the two approaches have reached an intersection. Researchers build computational cognitive models according to theories from linguistics, information science, and psychology, and examine their accuracy in predicting naturalistic data gathered from experimental linguistic and neurological studies.

Left corner parsing is a widely acknowledged parsing strategy that is proved to have psychological and neurological plausibility (e.g. Sugimoto et al., 2023; Yoshida, Noji, & Oseki, 2021). Various psycholinguistic theories (e.g. Oh, Clark, & Schuler, 2021; Rasmussen & Schuler, 2018) are structured based on this strategy. Left corner parsing frames sentence comprehension as a process under a pushdown store implementation of working memory (Rosenkrantz & Lewis, 1970). A hierarchical structure representation of sentences is formed during parsing. Working memory in this framework operates to (1) push the new derivation fragments to the store, (2) retrieve and compose derivation fragments from the store, and (3) maintain incomplete derivation fragments in the store (Shain, Blank, Fedorenko, Gibson, & Schuler, 2022). This process assigns a lexical label for each word and introduces it into complete fragments according to the grammar rules.

Left corner parsing has also proven its superiority in algorithmic robustness. Top-down and bottom-up parsing are two common that suggest establishing the analysis either from the highest level of the hierarchical structure to specific words or on the contrary. Both algorithms have drawbacks when being applied independently, which can be avoided by combining them together. Left corner parsing achieves this synthesis by making both lexical and grammatical decisions when processing each word.

In this paper, we aim to find out (1) to what extent left corner theories be proved as psychological and neurological plausible, (2) what challenges these theories encounter and (3) how they overcome these challenges.

Method

To prove that their theories are psycholinguistically plausible, the researchers build computational cognitive models according to the theoretical frameworks. Then by conducting correlation analyses between the results predicted by the models and the naturalistic data, they examine the robustness of their models and therefore the underlying theories.

The implementation of the computational models varies by the theories they are meant to prove. Some models map the lexical and syntactic information to the random vectors to enable a generalizable conclusion to be yielded from limited simulations (Rasmussen & Schuler, 2018), while others use vector representations pretrained from neural networks to make predictions comparable to real-world data (e.g. Sugimoto et al., 2023). They also vary by the ways that linguistic constraints are acquired. The constraints can be predefined by human researchers (e.g. Rasmussen & Schuler, 2018), be acquired through learning from massive data by models themselves (Sugimoto et al., 2023; Yoshida et al., 2021), or be set up by both manual input and machine learning (Oh et al., 2021).

The left corner parser proposed by Rasmussen and Schuler (2018) incrementally reads each word from the input sentence, represents the word with a randomly initialized vector, and then assigns a plausible category label according to the preterminal rules that associate words with syntactic constituents. A comparison between the label predicted from the built-up sentence structure and the incoming observation decides where the word is situated in the hierarchical structure. The model stores the lexical and syntactic information in the distributed associative memory (Anderson, Silverstein, Ritz, & Jones, 1977; Howard & Kahana, 2002; McClelland, McNaughton, & O'Reilly, 1995). Particularly, the observed word and its meaning, which are both represented by vectors, are associated with the outer product of these vectors. This means of information retrieval allows for the parallel processing of multiple analyses, as different analyses can be superposed in the association matrix weighted according to their probability. This implementation aligns more with the human memory system, which is verified by multiple neurological studies (Hasselmo & Wyble, 1997; Marr, Willshaw, & McNaughton, 1991; Treves & Rolls, 1994).

The model used in Oh et al. (2021) and Oh, Clark, and Schuler (2022) is established on a combination of neural networks and linguistic constraints. The lexical label for each word is drawn from a corpus re-annotated with the generalized categorial grammar (Nguyen, Van Schijndel, & Schuler, 2012). The model also calculates the logical predicates between nodes and represents them with associated nodes in the tree structure. Moreover, the built-in morphological rules associate transformations in lexical orthography with transformations between syntactic categories of words. These attributes guarantee the generalizability of the model allowing for the estimate of reading time in a large-scale dataset.

The left-corner models implemented in Sugimoto et al. (2023) and Yoshida et al. (2021) are based on recurrent neural network grammars (Dyer, Kuncoro, Ballesteros, & Smith, 2016). This neural network simultaneously analyzes the category label of words and the syntactic relationship within a hierarchical structure among them. Unlike the previous models, which are explicitly provided with constraints on grammar rules, the recurrent neural network grammars are trained on massive text data, and the grammar rules are boiled down from this training process.

Results

Using their left corner models, the researchers conduct simulations with sentences designed manually or drawn from corpora. The psycholinguistic plausibility of the left corner parsing strategy is demonstrated by the statistical analysis of the simulation results.

Rasmussen and Schuler (2018) manipulates the center-embedding constructions of the sentences. The results indicate that the parser simulates accurately the processing difficulty brought by the center-embedding structure. The model also reveals that the parsing failure is attributed to the increased memory demand in that each center embedding structure don't merge into the previously built-up analysis.

The structural processing model proposed in Oh et al. (2021) and Oh et al. (2022) demonstrates a better simulation of the reading time delay in self-paced reading, eye-tracking, and fMRI data than those large-scale language models trained with much more data. This result reflects the precedence of constraints over the unpredictable training in simulating psycholinguistic phenomena, indicating that the left corner parsing is more germane to human processing than the neural networks, which is motivated by the efficacy under the standard of engineering. The Oh et al. (2022) further claims that the transformer-based language models' bad performance is due to their superhuman accuracy in predicting the surprisal of rare words.

The left corner neural networks in Sugimoto et al. (2023) and Yoshida et al. (2021)

manifest a superiority to the sequential network and the hierarchical network implemented by top-down parsing strategy in predicting human brain activity. The results support that the hierarchical structure and left-corner parsing are more psychologically plausible in human sentence processing. Sugimoto et al. (2023) extends that syntactic composition is also stored in the memory system in addition to word predictability, given that expectation (i.e. surprisal) can't independently account for the psycholinguistic phenomena.

Discussion

In addition to the left-corner parsing theories, there are other competitive theories that are broad coverage (i.e. sufficiently articulated to predict word-by-word memory demand in arbitrary utterances), such as dependency locality theory (*abbr.* DLT, Gibson et al., 2000) and ACT-R (Adaptive Control of Thought-Rational) sentence processing theories (Lewis & Vasishth, 2013). These theories are generally referred to as memory-based theories since they propose the incremental processing demand to be attributed to building the structure for the unfolding sentence in the working memory.

A main challenge faced by this approach is from the frequency-based methods (Hale, 2001; Levy, 2008), which propose that cognitive loads are determined by the information contributed by a word toward the interpretation, instead of the working memory operations claimed by memory-based theories. This expectation-based approach is supported by empirical studies, which demonstrate that working memory effects doesn't emerge when surprisal is controlled (e.g. Demberg & Keller, 2008; Shain & Schuler, 2021; Van Schijndel & Schuler, 2013).

The left corner parser implemented in (Rasmussen & Schuler, 2018) reaches a unification of these two approaches. First, it builds the word-by-word structure during parsing and accounts for the memory effects of this structure building demand. Second, the parser also accounts for the derivation of surprisal as the result of renormalizing the reduced vectors after being resolved from the superposed states, incorporating probabilistic interpretation into the reasons for processing difficulty. This design contributes to a complete picture of sentence processing, being consistent with the claim by Levy and Gibson (2013) that neither expectation-based nor memory-based theories can independently account for the frequency and memory effects in sentence processing.

Conclusions

This paper probes into the research of sentence processing, reviewing five empirical studies about left corner parsing. We introduce the implementation of models proposed in these studies and their simulation experiments. The results demonstrate the psychological and neurological plausibility of left corner parsing in human sentence processing. However, it can not independently account for the frequency and memory effects, as proposed and proved by exponents of frequency-based theories. The model proposed by Rasmussen and Schuler (2018) incorporates both approaches into the parsing process, achieving both high simulation accuracy and remarkable psycholinguistic implications.

References

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological review*, *84*(5), 413.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210.

Dyer, C., Kuncoro, A., Ballesteros, M., & Smith, N. A. (2016, June). Recurrent neural network grammars. In K. Knight, A. Nenkova, & O. Rambow (Eds.), *Proceedings of the 2016 conference of the north American chapter of the association for computational linguistics: Human language technologies* (pp. 199–209). San Diego, California: Association for Computational Linguistics. Retrieved from `https://aclanthology.org/N16-1024` doi: 10.18653/v1/N16-1024

Gibson, E., et al. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, *2000*, 95–126.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Second meeting of the north american chapter of the association for computational linguistics.*

Hasselmo, M. E., & Wyble, B. P. (1997). Free recall and recognition in a network model of the hippocampus: simulating effects of scopolamine on human memory function. *Behavioural brain research*, *89*(1-2), 1–34.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of mathematical psychology*, *46*(3), 269–299.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Levy, R., & Gibson, E. (2013). *Surprisal, the pdc, and the primary locus of processing difficulty in relative clauses* (Vol. 4). Frontiers Media SA.

Lewis, R. L., & Vasishth, S. (2013). An activation-based model of sentence processing as skilled memory retrieval. In *Dictionary of world philosophy* (pp. 375–419). Routledge.

Marr, D., Willshaw, D., & McNaughton, B. (1991). *Simple memory: a theory for archicortex.* Springer.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological* *review*, *102*(3), 419.

Nguyen, L., Van Schijndel, M., & Schuler, W. (2012). Accurate unbounded dependency recovery using generalized categorial grammars. In *Proceedings of coling 2012* (pp. 2125–2140).

Oh, B.-D., Clark, C., & Schuler, W. (2021). Surprisal estimators for human reading times need character models. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers)* (pp. 3746–3757).

Oh, B.-D., Clark, C., & Schuler, W. (2022). Comparison of structural parsers and neural language models as surprisal estimators. *Frontiers in Artificial Intelligence*, *5*, 777963.

Rasmussen, N. E., & Schuler, W. (2018). Left-corner parsing with distributed associative memory produces surprisal and locality effects. *Cognitive science*, *42*, 1009–1042.

Rosenkrantz, D. J., & Lewis, P. M. (1970). Deterministic left corner parsing. In *11th annual symposium on switching and automata theory (swat 1970)* (pp. 139–152).

Shain, C., Blank, I. A., Fedorenko, E., Gibson, E., & Schuler, W. (2022). Robust effects of working memory demand during naturalistic language comprehension in language-selective cortex. *Journal of Neuroscience*, *42*(39), 7412–7430.

Shain, C., & Schuler, W. (2021). Continuous-time deconvolutional regression for psycholinguistic modeling. *Cognition*, *215*, 104735.

Sugimoto, Y., Yoshida, R., Jeong, H., Koizumi, M., Brennan, J. R., & Oseki, Y. (2023). Localizing syntactic composition with left-corner recurrent neural network grammars. *Neurobiology of Language*, 1–24.

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*(3), 374–391.

Van Schijndel, M., & Schuler, W. (2013). An analysis of frequency-and memory-based processing costs. In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies* (pp. 95–105).

Yoshida, R., Noji, H., & Oseki, Y. (2021). Modeling human sentence processing with left-corner recurrent neural network grammars. *arXiv preprint arXiv:2109.04939*.