Retrieving structures from memory causes difficulty during incremental processing

Cory Shain[1], Marten van Schijndel[1], Richard Futrell[2], Edward Gibson[2], and William Schuler[1]
[1]OSU,[2]MIT

Any incremental model of sentence processing where an abstract meaning representation is built up word by word must involve storage and retrieval of information about previously encountered material from some memory store. The retrieval operations have been hypothesized to be associated with increased processing time [4, 8, 12], and this prediction has been borne out in experiments using constructed stimuli [4, 5, 1, 11]. However, memory-based latency effects have been null or even negative in broad-coverage reading time experiments using naturally-occurring text data that included baseline controls for *n*-gram and PCFG surprisal [2, 10]. There are potential confounds associated with each type of stimulus. The effects observed in constructed stimuli could be due to (1) information theoretic phenomena (e.g., surprisal) that such experiments rarely control for, (2) limited syntactic domains (e.g., relative clauses), or (3) lack of discourse context, rather than due to difficulty retrieving information from working memory. On the other hand, the lack of positive latency effects in studies using naturally-occurring input could be (1) because of the small number of subjects – ten – in the Dundee corpus [7] used by e.g. [2, 10] or (2) because naturally-occurring newswire texts might contain too low a proportion of memory-intensive constructions to reveal a generalized memory effect.

The present work seeks to address these concerns using the new Natural Stories corpus [3], a 'constructed-natural' self-paced reading corpus that occupies an intermediary position between constructed stimuli on the one hand and naturally-occurring stimuli on the other. The corpus consists of ten fluent, context-rich narrative texts that are constructed in order to contain a high proportion of memory-intensive constructions, with self-paced reading data collected from 181 subjects, yielding 848,207 reading events. These data were partitioned into exploratory (1/3) and confirmatory (2/3) datasets, and the exploratory corpus was used for empirical selection from a number of plausible implementations of two broad models of human sentence processing: the Dependency Locality Theory (DLT) [4] and left-corner parsing [6, 8]. The DLT predicts processing difficulty as a function of dependency length (defined as the number of intervening discourse referents), while left-corner models predict processing difficulty as a function of parser operations that involve memory retrieval.

Using linear mixed-effects models and a strong baseline that includes controls for both *n*-gram and PCFG surprisal [9], we find highly significant inhibitory effects for both DLT integration cost and a left-corner parsing operation that corresponds to the termination of a phrasal constituent (Table 1). The DLT effect improves considerably with certain independently-motivated modifications to its implementation (see [9] for details). The DLT and left-corner predictors make independent contributions to model fit, with both predictors improving significantly over the baseline and over each other individually. To our knowledge, this is the first strong evidence of memory effects in broad-coverage sentence processing, and our results suggest the cognitive reality of important constructs from syntactic theory such as dependency (DLT) and constituency (left-corner parsing). The existence of separate effects for the DLT, which encodes semantic information such as asymmetries in referential status, and left-corner models, which are sensitive only to syntactic tree configurations, might suggest the existence of separate effects for computing dependencies on the one hand and retrieving syntactic structures from memory on the other, a question we leave to future research. We also note that all effect sizes are quite small ($\leq$ 4 ms), a fact that might shed light on the aforementioned discrepancy between experiments based on naturally-occurring vs. constructed stimuli. Our results suggest that memory effects do indeed exist in broad-coverage sentence processing, but are so small that they may not be observable without difficult stimuli and/or large corpora that include data from many subjects.

	$\hat{\beta}$ (ms)	*t*-value	*p*-value
DLT	0.466	4.50	1.11e-05
DLT-CM (modified DLT)	1.13	6.48	4.87e-10
Right edge of constituent	3.88	8.15	2.33e-14

Table 1: Results from DLT and left-corner predictors show robust memory effects on reading times

Acknowledgements:
This work was supported by grants from the National Science Foundation: Graduate Research Fellowship Program Award DGE-1343012 to MvS; Doctoral Dissertation Research Improvement Award 1551543 to RF; Linguistics Program Award 1534318 to EG.

References

[1] Marisa Ferrara Boston, John T. Hale, Shravan Vasishth, and Reinhold Kliegl. Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, 26(3):301–349, 2011.

[2] Vera Demberg and Frank Keller. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2):193–210, 2008.

[3] Richard Futrell, Edward Gibson, Hal Tily, Anastasia Vishnevetsky, Steve Piantadosi, and Evelina Fedorenko. Natural stories corpus. in prep.

[4] Edward Gibson. The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, language, brain: Papers from the first mind articulation project symposium*, pages 95–126, Cambridge, MA, 2000. MIT Press.

[5] Daniel J. Grodner and Edward Gibson. Consequences of the serial nature of linguistic input. *Cognitive Science*, 29:261–291, 2005.

[6] Philip N. Johnson-Laird. *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press, Cambridge, MA, USA, 1983.

[7] Alan Kennedy, James Pynte, and Robin Hill. The Dundee corpus. In *Proceedings of the 12th European conference on eye movement*, 2003.

[8] Richard L. Lewis and Shravan Vasishth. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3):375–419, 2005.

[9] Cory Shain, Marten van Schijndel, Richard Futrell, Edward Gibson, and William Schuler. Memory access during incremental sentence processing causes reading time latency. In *Proceedings of the Computational Linguistics for Linguistic Complexity Workshop*, pages 49–58. Association for Computational Linguistics, 2016.

[10] Marten van Schijndel, Luan Nguyen, and William Schuler. An analysis of memory-based processing costs using incremental deep syntactic dependency parsing. In *Proc. of CMCL 2013*. Association for Computational Linguistics, 2013.

[11] Titus von der Malsburg, Reinhold Kliegl, and Shravan Vasishth. Determinants of scanpath regularity in reading. *Cognitive Science*, 2015.

[12] Stephen Wu, Asaf Bachrach, Carlos Cardenas, and William Schuler. Complexity metrics in an incremental right-corner parser. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL'10)*, pages 1189–1198, 2010.