Coreference and Discourse Focus in Broad-Coverage Stimuli

Evan Jaffe, Cory Shain, William Schuler

jaffe.59@osu.edu

The Ohio State University

Overview

- Prominence facilitates anaphor resolution [1, 6]
- **Problem:** Focus, as defined by syntactic clefts and observed in human reading times of constructed stimuli, is potentially confounded with frequency
- **Observation:** Results from naturally-occurring, contextualized stimuli can complement results from constructed stimuli [4]
- **Question:** Do focus effects on reading times observed using constructed stimuli generalize to naturalistic stimuli?
- **Approach:** Predict self-paced reading times using coreference-based predictors designed to generalize focus definition to naturalistic stimuli

Background

- Multiple syntactic, semantic, pragmatic, and discourse factors affect referent focus when modeling anaphor resolution [10]
- Prior context can effectively model focus [14]
- *Thematization* [13] defined as count of referent occurrences in a discourse, and used to focus a referent prior to presentation of a target stimulus

Data

- Natural Stories Corpus [7] consists of 10 constructed-natural stories, with self-paced reading (SPR) times for 181 subjects
- Stories augmented with more memory-taxing syntactic constructions, rare lexical items, and idioms than in naturally-occurring text
- Total items: 768,023 from 485 sentences
- Data partitioned into exploratory (1/3) and confirmatory (2/3) subsets
- Coreference annotation largely follows OntoNotes guidelines [16], but adds anaphoric determiners like *its, hers* etc.

Methods

- Ablative likelihood ratio testing of linear mixed effects models [2] using by-subject random slopes and by-subject random intercepts for all predictors
- Filter RTs less than 100ms, exceeding 3000 ms, or exceeding 2 standard deviations from mean
- All predictors z-transformed
- Box-Cox [3] power transform of reading times to match LMER assumptions of normality

Coreference Annotation Example

The Lord saw the severity of the problem the people faced and suggested a contest could solve the problem.
He said that whoever could kill the boar and bring as proof its head ... would be rewarded with land and fame.
It was the people of Bradford ... who rejoiced at this proclamation but one question remained: who would kill the boar?

Predictors

- Baseline
  - *Word Length* - measured in characters
  - *Syntactic Surprisal* - Probabilistic Context-Free Grammar (PCFG) surprisal estimate using an incremental parser [15] over Generalized Categorial Grammar [12]
  - *N-gram Suprisal* - 5-gram surprisal using KenLM [11] over GigaWord corpus [9]
  - *Story Position* - percent completion of story, scaled to [0,1]
- Predictors of Interest
  - *Mention Count* - running total count of mentions for the referent
  - *Antecedent Distance Word* - distance to most recent mention measured in words [8]
  - *Antecedent Distance Referent* - distance to most recent mention measured in referents, operationalized as nouns or verbs
- Spillover
  - To account for delay in the time-course of processing, spill over position [5] was optimized for predictors using ordinary least squares on exploratory data
  - Mention Count and Syntactic Suprisal were stronger predictors in exploratory data when spilled over by one word position, and were selected for confirmatory tests

Example Predictor Values

	The	Lord_i	saw	the	severity	of	the	problem_j	the	people	faced	and	suggested	a	contest	could	solve	the	problem_j.
MentionCount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**
WordDistance	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**
ReferentDistance	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**5**
	He_i	said	that	whoever	could	kill	the	boar_k	and	bring	as	proof	its_k	head	would	be	rewarded	with	land	and	fame.
MentionCount	**1**	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0
WordDistance	**18**	0	0	0	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0	0	0
ReferentDistance	**9**	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	0	0	0	0	0	0

Results

	Effect Size (ms)	
Effect	Predictor units	Z
Word Length	2.17	4.23
Syntactic Surprisal	0.36	1.65
5-gram Surprisal	2.34	3.57
Story Position	-19.2	-6.62
MentionCount***	-0.14	-2.81

Table 1: Effect sizes for baseline and predictors of interest on confirmatory partition of data. Mention Count is highly significant ($p = 7.05e-5$). Negative effect direction indicates a speed-up in reading times. Effect estimates in milliseconds are backtransformed from Box-Cox estimates and only valid at the backtransformed mean, holding all other effects at their means. Z shows β-effect in milliseconds per unit of standard deviation. Predictor Units are the effect size in milliseconds, rescaled to the original predictors' units. Model includes observations from spilled over anaphors (proforms and fully referring expressions), totaling 59,632 observations. Word Length is measured in characters, Surprisal is measured in bits, and Story Position is the proportion of sentences completed, scaled between 0 and 1. Note that Mention Count ranges from 1-90, so a word referring to an entity with 70 previous mentions is predicted to be read approximately 10ms faster, relative to a singleton mention.

Conclusion

- **Focus facilitation effect for broad-coverage stimuli observed for coreference-based measure of discourse prominence, Mention Count**
- Antecedent distance-based predictors were not significant on exploratory partition and not run on confirmatory
- Strong effect of Story Position evidence of importance of controlling for order effects in SPR

References

[1] Amit Almor. Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review*, 106(4):748–765, October 1999.

[2] Douglas Bates, Martin Maechler, and Bin Dai. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-31, 2008.

[3] George E. P. Box and David R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):211–252, 1964.

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

[5] Kate Erlich and Keith Rayner. Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning & Verbal Behavior*, 22:75–87, 1983.

[6] Stephani Foraker and Brian McElree. The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, 56(3):357–383, 2007.

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

[8] 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.

[9] David Graff and Christopher Cieri. *English Gigaword LDC2003T05*, 2003.

[10] Steven B. Greene, Gail McKoon, and Roger Ratcliff. Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18:266–283, 1992.

[11] Kenneth Heafield, Ivan Pouzyrevsky, Jonathan H. Clark, and Philipp Koehn. Scalable modified Kneser-Ney language model estimation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 690–696, 2013.

[12] Luan Nguyen, Marten van Schijndel, and William Schuler. Accurate unbounded dependency recovery using generalized categorial grammars. In *Proceedings of COLING 2012*, pages 2125–2140, Mumbai, India, 2012.

[13] Charles A. Perfetti and Susan R. Goldman. Thematization and sentence retrieval. *Journal of Verbal Learning and Verbal Behavior*, 13(1):70 – 79, 1974.

[14] Antje Sauermann, Ruth Filik, and Kevin B. Paterson. Processing contextual and lexical cues to focus: Evidence from eye movements in reading. *Language and Cognitive Processes*, 28(6):875–903, 2013.

[15] Marten van Schijndel and William Schuler. An analysis of frequency- and memory-based processing costs. In *Proceedings of NAACL-HLT 2013*. Association for Computational Linguistics, 2013.

[16] R. Weischedel, M. Palmer, M. Marcus, E. Hovy, S. Pradhan, L. Ramshaw, N. Xue, A. Taylor, J. Kaufman, M. Franchini, El-Bachouti M., Belvin R., and A. Houston. Ontonotes release 5.0. `https://catalog.ldc.upenn.edu/ldc2013t19`, 2013. LDC Catalog No.: LDC2013T19.

Acknowledgements

- This work was supported by NSF GRFP grant no. DGE-1343012, and NSF grant no. 1551313. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.