Coreference and Focus in Reading Times

Evan Jaffe, Cory Shain, William Schuler
Cognitive Modeling and Computational Linguistics (CMCL) 2018
Salt Lake City, UT
1/7/2018

Background

Linguistic focus improves/facilitates coreference resolution (Foraker, McElree 2007; Almor 1999)

Constructed stimuli using syntactic clefts:

It was the robin that ate the fruit./What the robin ate was the fruit.

The bird seemed quite satisfied.

Focused word recalled better

Motivation

Syntactic clefts are low frequency; potential confounds of frequency or oddball effect

Big Picture: Do processing effects using constructed stimuli generalize? Are they reproducible using naturalistic stimuli?

Motivation

Syntactic clefts are low frequency; potential confounds of frequency or oddball effect

Big Picture: Do processing effects using constructed stimuli generalize? Are they reproducible using naturalistic stimuli?

No, constructed stimuli different

Frank and Bod (2011) hierarchical vs. linear models no significant difference on Dundee

Van Schijndel et al. (2013) facilitation rather than expected cost for memory-intensive retrieval on Dundee

Motivation

Syntactic clefts are low frequency; potential confounds of frequency or oddball effect

Big Picture: Do processing effects using constructed stimuli generalize? Are they reproducible using naturalistic stimuli?

No, constructed stimuli different

Frank and Bod (2011) hierarchical vs. linear models no significant difference on Dundee

Van Schijndel et al. (2013) facilitation rather than expected cost for memory-intensive retrieval on Dundee

Yes, reproduce

Shain et al. (2016) predicted inhibitory effect of syntactic dependency length on Natural Stories

Brennan et al. (2016) hierarchical grammars predict time course (fMRI) on naturalistic stimuli

Question Definition

Do linguistic focus effects generalize to broad-coverage naturalistic stimuli?

Must redefine linguistic focus for naturalistic stimuli without clefts

Use coreference as a measure of linguistic focus

- less frequency confound
- prevalent in many genres
- similar to existing coreference-based measures of focus

Coreference-based Focus Predictors

Distance

- Givon 1983 - leftward distance in clauses between anaphor and antecedent
- DLT (Gibson 2000) - Distance between governor and dependent affects processing ease (not focus per se)
- This work uses intervening word and referent-based distance measures

Coreference-based Focus Predictors

Distance

- Givon 1983 - leftward distance in clauses between anaphor and antecedent
- DLT (Gibson 2000) - Distance between governor and dependent affects processing ease (not focus per se)
- This work uses intervening word and referent-based distance measures

Topicality

- Topicality in Discourse (Givon 1983) - *Persistence*: number of uninterrupted clauses to the right that an entity continues as a semantic argument
- *Thematization* (Perfetti Goldman 1973) - Total count of entity mentions
- This work generalizes the measure to a running count in order to deal with incremental processing

Data

Natural Stories Corpus (Futrell et al, in prep)

10 stories, 181 participants

Self-paced reading paradigm (SPR)

768,023 events after filtering outliers and inattentive subjects (59,632 anaphor events)

designed to include some memory intensive constructions including topicalization, clefting, idioms, etc., striking a balance between constructed and natural stimuli

Methods

Linear Mixed Effects Regression (LMER) models

Likelihood Ratio Test: baseline model vs. baseline+main predictor model

Dependent variable: Reading times

All predictors centered and z-transformed prior to model fitting

Baseline Predictors

Word Length - in characters

N-gram Surprisal - 5-gram over Gigaword (Graff and Cieri 2003) using KENLM (Heafield et al. 2013)

$$S(w_i) = -\,log\ P(w_i | w_{i-n} ... w_{i-1})$$

Syntactic Surprisal - PCFG using incremental parser over generalized categorial grammar (van Schijndel 2016)

$$S(w_i) = -\,log\ P(T_i = w_i | T_1 ... T_{i-1} = w_1 ... w_{i-1})$$

Story Position - proportional sentence location in narrative, intended to model order effects of task learning or fatigue

Main Predictors

Distance

- Coreference Length Word - distance from anaphor to antecedent measured by intervening words
- Coreference Length Referent - distance from anaphor to antecedent measured by intervening referents (nouns or verbs)

Topicality

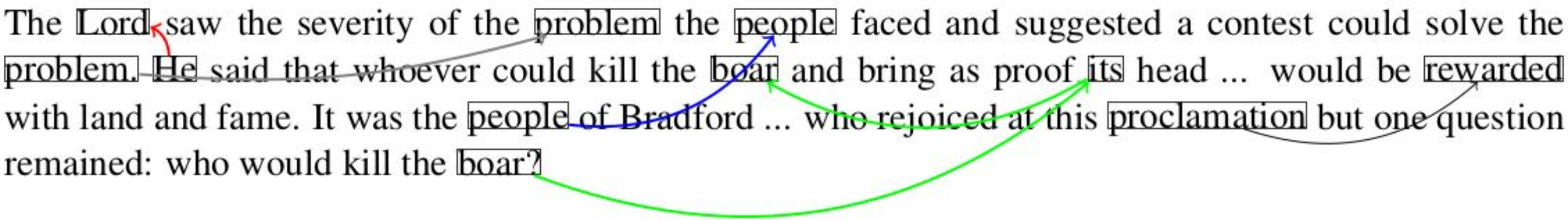
- Mention Count - running count of mentions for a given entity

Coreference Annotation

Natural Stories corpus augmented with identity coreference annotation largely following OntoNotes 5.0 (Weischedel et al. 2013) guidelines

Pronouns, verbs, nouns can be marked as anaphors

Also added possessives (*his*, *her*, *its*, ...)



	The	Lord$_i$	saw	...	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	**1**
WordDistance	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	**5**

	He$_i$	said	that	whoever	could	kill	the	boar$_k$	and	bring	as	proof	its$_k$	head	would	...
MentionCount	**1**	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	...
WordDistance	**18**	0	0	0	0	0	0	0	0	0	0	0	**4**	0	0	...
ReferentDistance	**9**	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	...

Anaphors can be fully referring or proforms

Distances can span beyond sentences

MentionCount increments each time the referent is mentioned

Box-Cox Power Transform

Reading time data was transformed to match assumptions of normality by LMER

Box-Cox (1964) equation:

$$y^{(\lambda)} = \frac{y^{\lambda} - 1}{\lambda}$$

where $\lambda \neq 0$

$\lambda = -0.63$ determined from built-in R function

Also done in Shain et al. (2016)

Spillover

Delays in time course of processing effects modeled using spillover (Erlich and Rayner 1983)

Effect of independent variable predicted to occur *n* words later

Baseline and main predictors best *n* optimized on exploratory data - MentionCount and PCFG surprisal strongest at spillover 1 (approximately 300ms, fits with syntactic processing time course)

Results

Facilitation for increased MentionCount

t-value: -4.085, *** (*p*=7.05e-05)

	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

Results

Facilitation for increased MentionCount

t-value: -4.085, *** (p=7.05e-05)

	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

Predicted inhibitory effect of increased word length, surprisal

Facilitation for increased story position

MentionCount predictor units vary from 0-90, roughly 10ms difference between large and small MentionCount

Discussion

Why no distance effects?

Demberg and Keller (2008) also do not show distance effects for syntactic dependencies (Dundee), except for certain parts of speech

Contrast with Shain et al. (2016), who do find inhibitory effect of dependency length for Natural Stories corpus

Discussion

Why no distance effects?

Demberg and Keller (2008) also do not show distance effects for syntactic dependencies (Dundee), except for certain parts of speech

Contrast with Shain et al. (2016), who do find inhibitory effect of dependency length for Natural Stories corpus

Dependencies limited to sentence length, whereas coreference can span entire stories

Lack of sufficiently strong effects for very long distance coreference could be masking a real effect for shorter coreference distances

Future work could limit to intrasentential coreference or cap distances to look for distance effects

Discussion

Story Position very strong predictor - recommend including order effect predictor for similar studies

Conclusion

Linguistic focus effects do generalize to naturalistic stimuli

MentionCount coreference-based predictor is a suitable measure of linguistic focus for naturalistic stimuli

Acknowledgements

Thank you to four anonymous reviewers.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under 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.

References

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

George E. P. Box and David R. Cox. 1964. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological) 26(2):211–252. http://www.jstor.org/stable/2984418.

Jonathan R. Brennan, Edward P. Stabler, Sarah E. Van Wagenen, Wen-Ming Luh, and John T. Hale. 2016. Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. Brain and Language 157:81 – 94. https://doi.org/10.1016/j.bandl.2016.04.008.

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

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

References

Stephani Foraker and Brian McElree. 2007. The role of prominence in pronoun resolution: Active versus passive representations. Journal of Memory and Language 56(3):357–383. https://doi.org/10.1016/j.jml.2006.07.004.

Stefan Frank and Rens Bod. 2011. Insensitivity of the human sentence-processing system to hierarchical structure. Psychological Science .

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

Talmy Givón. 1983. Topic continuity in discourse: An introduction. In Talmy Givón, editor, Topic Continuity in Discourse: A Quantitative Cross-Language Study, John Benjamins, Amsterdam, pages 1–41.

David Graff and Christopher Cieri. 2003. English Gigaword LDC2003T05.

References

Steven B. Greene, Gail McKoon, and Roger Rat-
cliff. 1992. Pronoun resolution and discourse mod-
els. Journal of Experimental Psychology: Learning, Memory, & Cognition 18:266–283.
Daniel J. Grodner and Edward Gibson. 2005. Consequences of the serial nature of linguistic input. Cognitive Science 29:261–291.
John Hale. 2001. A probabilistic earley parser as a psycholinguistic model. In Proceedings of the second meeting of the North American chapter of the Association for Computational Linguistics. Pittsburgh, PA, pages 159–166.
Kenneth Heafield, Ivan Pouzyrevsky, Jonathan H.
Clark, and Philipp Koehn. 2013. Scalable modified Kneser-Ney language model estimation. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. Sofia, Bulgaria, pages 690–696.

References

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

Alessandro Lopopolo, Stefan L. Frank, Antal van den Bosch, and Roel M. Willems. 2017. Using stochastic language models (slm) to map lexical, syntactic, and phonological information processing in the brain. PLOS ONE 12(5):1–18. https://doi.org/10.1371/journal.pone.0177794.

Brian McElree. 2001. Working memory and focal attention. Journal of Experimental Psychology, Learning Memory and Cognition 27(3):817–835.

Irene Fernandez Monsalve, Stefan L. Frank, and Gabriella Vigliocco. 2012. Lexical surprisal as a general predictor of reading time. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, Stroudsburg, PA, USA, EACL '12, pages 398–408. http://dl.acm.org/citation.cfm?id=2380816.2380866.

References

Luan Nguyen, Marten van Schijndel, and William Schuler. 2012. Accurate unbounded dependency recovery using generalized categorial grammars. In Proceedings of the 24th International Conference on Computational Linguistics (COLING '12). Mumbai, India, pages 2125–2140.

Charles A. Perfetti and Susan R. Goldman. 1974. Thematization and sentence retrieval. Journal of Verbal Learning and Verbal Behavior 13(1):70 – 79. https://doi.org/http://dx.doi.org/10.1016/S0022-5371(74)80032-0.

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.

Antje Sauermann, Ruth Filik, and Kevin B. Paterson. 2013. Processing contextual and lexical cues to focus: Evidence from eye movements in reading. Language and Cognitive Processes 28(6):875–903. https://doi.org/10.1080/01690965.2012.668197.

References

Cory Shain, Marten van Schijndel, Richard Futrell, Edward Gibson, and William Schuler. 2016. Memory access during incremental sentence processing causes reading time latency. COLING 2016, workshop on Computational Linguistics for Linguistic Complexity.

Nathaniel J. Smith and Roger Levy. 2013. The effect of word predictability on reading time is logarithmic. Cognition 128:302–319.

Patrick Sturt and Vincent Lombardo. 2005. Processing coordinate structures: Incrementality and connectedness. Cognitive Science 29:291–305.

References

Marten van Schijndel, Andy Exley, and William Schuler. 2013. A model of language processing as hierarchic sequential prediction. Topics in Cognitive Science 5(3):522–540.
Marten van Schijndel and William Schuler. 2015. Hierarchic syntax improves reading time prediction. In Proceedings of NAACL-HLT 2015. Association for Computational Linguistics.
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. 2013. Ontonotes release 5.0. https://catalog.ldc.upenn.edu/ ldc2013t19. LDC Catalog No.: LDC2013T19.
Wayne Wicklegren. 1977. Speed-accuracy tradeoff and information processing dynamics. Acta Psychologica 41:67–85.

Reverse Box-Cox Power Transform

$$y_t = \begin{cases} \exp(w_t) & \lambda = 0; \\ (\lambda w_t + 1)^{1/\lambda} & \text{otherwise.} \end{cases}$$