What Does Coreference Cost? Regressing Reading Times to Coreference-based Predictors

Evan Jaffe, Cory Shain, William Schuler
The Ohio State University
7/1/17
Midwest Cognitive Science Conference
Miami University

What Is Coreference?

When the same entity is referred to multiple times

E.g., "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?"

Coreference Resolution is the binding of the *anaphor* back to its *antecedent*

Implications for Memory and Retrieval

Coreference resolution assumed to incur processing costs due to memory access, similar to syntactic binding problem
(Felser, Phillips, and Wagers 2017)

Processing costs proportional to intervening material
(e.g., Dependency Locality Theory; Gibson 2000)

Are reading times proportional to antecedent decay? No...

Implications for Memory and Retrieval

This work: 4 distance-based measures **NOT** significant predictors of reading times

Lack of distance effects raises issues for any model that includes decay
(e.g., DLT; Gibson 2000, ACT-R; Anderson 1996)

What about focus?

Coreference and Focal Attention

"Subset of representations in WM ... perhaps just one" (McElree 01)

Pronouns generally produced if referent is predictable/salient/accessible

Processing load reduced when antecedent is more prominent (Nicol & Swinney 2003)

Coreference resolution speed or accuracy can tell us about memory access, lexical retrieval and binding

Coreference and Focal Attention

Evidence for distinct psychological state/process for retrieving focused items

Probe recognition speeds 30-50% higher when no activity intervenes between study and test (McElree 1996, 1998)

Actively maintained representations (Gundel 1999)

Speed-accuracy tradeoff (SAT) task measures multiple acceptability judgements during sentence, has shown focus effect as improved accuracy, but not speed (Foraker & McElree 2007)

Decay or Focus?

Is decay/distance to antecedent predictive of reading time? No...

Is focus of antecedent predictive of reading time? Yes!

Decay or Focus?

Is decay/distance to antecedent predictive of reading time? No...

Is focus of antecedent predictive of reading time? Yes!

Self-paced reading time (SPR) corpus, annotated with coreference

Linear Mixed Effects Regression

Likelihood Ratio Test

Main Results: Sneak Preview

Focus, as measured by **coreference chain size**, shows a significant negative effect on reading time:

I.e., a word is read more quickly if its referent is mentioned more times

Distance to antecedent was **NOT** predictive of reading time

Natural Stories

Corpus of self-paced reading times (Futrell, Gibson, Tily, Vishnevetsky, Piantadosi, & Fedorenko, in prep)

180 subjects, 10 short stories

768,023 reading events (after boundary items removed)

Data split: ⅓ dev, ⅔ test - allows predictor choice and obviates multiple trials corr. on test

Natural Stories

Broad-coverage, naturalistic stimuli - balance between constructed and natural stimuli

Adds some memory-intensive constructions like clefting and idioms

Avoids potential oddball effect from constructed stimuli used in previous studies

Coreference Annotation and Predictors

"The Lord$_0$ saw the severity of the problem$_0$ the people$_0$ faced and suggested a contest could solve the problem$_1$. He$_1$ said that whoever could kill the boar$_0$ and bring as proof its$_1$ head ... would be rewarded$_0$ with land and fame. It was the people$_1$ of Bradford ... who rejoiced at this proclamation$_1$ but one question remained: who would kill the boar$_2$?"

Each word annotated with its most recent antecedent, forming a chain

Corefsize - measure of focus that is the cumulative number of times the referent has been mentioned up to that point

Coreference Annotation and Predictors

"The Lord$_0$ saw the severity of the problem$_0$ the people$_0$ faced and suggested a contest could solve the problem$_{11}$. He$_{18}$ said that whoever could kill the boar$_0$ and bring as proof its$_5$ head ... would be rewarded$_0$ with land and fame. It was the people$_{38}$ of Bradford ... who rejoiced at this proclamation$_{21}$ but one question remained: who would kill the boar$_{38}$?"

Coreflenw - distance from anaphor to antecedent in intervening words (shown)

Coreflenwlog - log transform of Coreflenw

Coreflenr - distance in intervening referents (nouns and verbs)

Coreflenrlog - log transform of Coreflenr

Baseline Predictors

Best baseline predictors chosen following work from Shain et al. (2016)

Word Length - longer words often read more slowly (Rayner and Pollatsek, 1987)

Sentence Position - earlier words often more informative and slower to read (boundary words removed)

Baseline Predictors

5-gram Forward Probability Surprisal - KENLM estimation; backoff-smoothed linear combination of n-grams (Monsalve et al., 2012; Van Schijndel and Schuler, 2015)

Syntactic Surprisal - probability of tree structure at given word (Fossum and Levy, 2012; Van Schindel and Schuler, 2015)

Surprisal: log(1/P(x))

Linear Mixed Effects Model

c.(((fdur^-0.626262626262626 - 1)/-0.626262626262626)) ~ z.(wlen) + z.(sentpos) + z.(fwprob5surp) + z.(totsurpS1) + z.(corefsizeS1) + (1 + z.(wlen) + z.(sentpos) + z.(fwprob5surp) + z.(totsurpS1) + z.(corefsizeS1) | subject) + (1 | word)

Box-Cox Transform (1964) of reading time variable to reduce model bias and normalize variance

Fixed effects included for all predictors, all subject random effects, and item random effects

Baseline model is same minus fixed corefsize effect

Spillover

Corefsize and syntactic surprisal are "spilled over" by one word index

Use reading time of subsequent word, assuming processing "spills over" (Ehrlich and Rayner 1983)

E.g., "The Lord saw the severity of the problem"

Reading time for "saw" would be used with corefsize predictor from "Lord"

Spillover position for all predictors optimized on dev/exploratory data partition

Coreference Size Speeds Reading

Test partition results:

	# Observations	CorefsizeS1 Effect estimate (ms/z)	P value
All Words	512,654	-3.25	4.40e-21***
Pronoun-filtered	27,923	-6.37	2.98e-13***

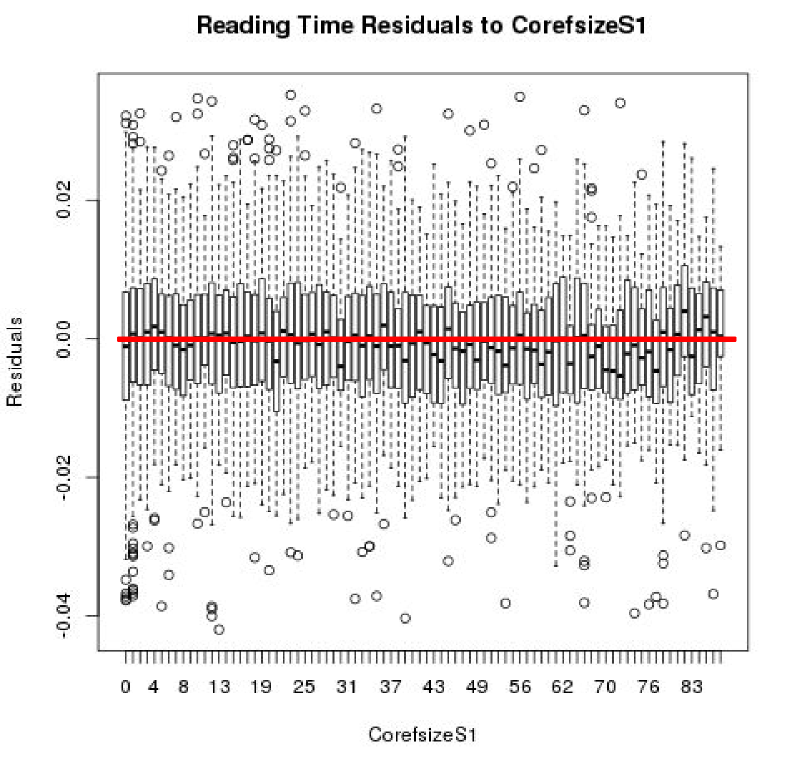
Negative trend, extrema not responsible

Linear baseline model (no random effects, no main effect)

Residuals show small negative effect of CorefsizeS1

Extrema included variety of mention types, ("she", "her", "herself"), syntactic positions (relative clause, matrix clause), sentence positions, S1 word length, etc.

No obvious confound



Discussion

Semantic priming effect as causing focus (e.g., Swinney et al. 1979)

Foraker (2007) accuracy effect (but not speed) vs. current speed effect could be explained by different task demands (SAT vs. SPR)

There might be accuracy/strength effect for distance not measured by SPR

Future work: Replication with eye-tracking, other definitions of focus/distance
Gundel (1999) differentiates psychological, semantic, and contrastive focus,
Givon (1983) measures distance in clauses

Conclusion

More focused referents are accessed more quickly, leading to faster self-paced reading times

No evidence of distance/decay-based speed changes

Supports focus effect - unique memory access for focused vs. not focused items

Consistent with Foraker (2007) focus effect, reproduced with broad-coverage naturalistic stimuli

Thanks!

Thank you to the OSU Linguistics Department, Cognitive and Computational Approaches to Linguistics (CaCL) lab group, David King, Lifeng Jin, Nathan Rasmussen

This material is based upon work supported by the NSF Graduate Research Fellowship Program under grant no. DGE-1343012

References

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. American Psychologist, 51(4), 355.
Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. Journal of the Royal Statistical Society, B, 26(211-234).
Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: eye movements and immediacy of processing. Journal of Verbal Learning and Verbal Behavior, 22, 75–87.
Felser, C., Phillips, C., & Wagers, M. (2017).Editorial: Encoding and Navigating Linguistic Representations in Memory. Frontiers in Psychology. 8(164)
Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. Journal of Memory and Language, 56(3), 357–383.
Fossum, V., & Levy, R. (2012). Sequential vs. hierarchical syntactic models of human incremental sentence processing. In Proceedings of CMCL 2012. Association for Computational Linguistics.
Futrell, Gibson, Tily, Vishnevetsky, Piantadosi, & Fedorenko, in prep.

References

Gibson, E. (2000). 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. MIT Press.

Givón, T.: (1983). Topic Continuity in Discourse. John Benjamins Publishing Company.

Gundel, J.K. (1999). On Different Kinds of Focus. Focus: Linguistic, cognitive, and computational perspectives. 293.Cambridge University Press.

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

McElree, B. (1996). Accessing short-term memory with seman- tic and phonological information: A time-course analysis. Memory & Cognition, 24, 173–187.

McElree, B. (1998). Attended and non-attended states in working memory: accessing categorized structures. Journal of Memory and Language, 38, 225–252.

References

Monsalve, I., Frank, S. L., & Vigliocco, G. (2012). Lexical surprisal as a general predictor of reading time. Eacl 2012, 398–408.

Nicol, J. L., & Swinney, D. A. (2003). The psycholinguistics of anaphora. In A. Barss (Ed.), Anaphora: A reference guide (pp. 72–104). Oxford, UK: Blackwell.

Rayner, K. & Pollatsek, A. 1987. Eye movements in reading: A tutorial review. In M. Coltheart, editor, Attention and performance XII: the psychology of reading., pages 327–362. Lawrence Erlbaum Associates, London, UK.

Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research

Shain, C., van Schijndel, M., Futrell, R., Gibson, E., & Schuler, W. (2016). Memory access during incremental sentence processing causes reading time latency. COLING 2016, Workshop on Computational Linguistics for Linguistic Complexity.

References

Swinney, D., Onifer, W. Prather, P. & Hirshkowitz M. (1979). Semantic facilitation across sensory modalities in the processing of individual words and sentences. Memory and Cognition, 7(3), 54-178.

Schijndel, M. Van. & Schuler, W. (2015). Hierarchic syntax improves reading time prediction. HLT-NAACL 2015 - Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings of the Main Conference, (2011), 1597–1605.