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:

\[ \text{It was the } \textit{robin} \text{ that ate the fruit.} / \text{What the robin ate was the } \textit{fruit}. \]

\[ \text{The bird seemed quite satisfied.} \]

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 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?

<table>
<thead>
<tr>
<th>MentionCount</th>
<th>The</th>
<th>Lord</th>
<th>saw</th>
<th>...</th>
<th>the</th>
<th>problem</th>
<th>the</th>
<th>people</th>
<th>faced</th>
<th>and</th>
<th>suggested</th>
<th>a</th>
<th>contest</th>
<th>could</th>
<th>solve</th>
<th>the</th>
<th>problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>He</th>
<th>said</th>
<th>that</th>
<th>whoever</th>
<th>could</th>
<th>kill</th>
<th>the</th>
<th>boar</th>
<th>and</th>
<th>bring</th>
<th>as</th>
<th>proof</th>
<th>its</th>
<th>head</th>
<th>would</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Effect</th>
<th>Effect Size (ms)</th>
<th>Predictor units</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Length</td>
<td>2.17</td>
<td></td>
<td>4.23</td>
</tr>
<tr>
<td>Syntactic Surprisal</td>
<td>0.36</td>
<td></td>
<td>1.65</td>
</tr>
<tr>
<td>5-gram Surprisal</td>
<td>2.34</td>
<td></td>
<td>3.57</td>
</tr>
<tr>
<td>Story Position</td>
<td>-19.2</td>
<td></td>
<td>-6.62</td>
</tr>
<tr>
<td>MentionCount***</td>
<td>-0.14</td>
<td></td>
<td>-2.81</td>
</tr>
</tbody>
</table>
Results

Facilitation for increased MentionCount

t-value: -4.085, *** \((p=7.05e^{-05})\)

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


References


References

References


References

References

Reverse Box-Cox Power Transform

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