Local Coherences in the Wild: An Eye-tracking Corpus Study

Klinton Bicknell[1], Vera Demberg[2], and Roger Levy[1]

[1]University of California, San Diego [2]University of Edinburgh

Abstract

The phenomenon of *local coherences* has recently generated significant interest in the sentence processing literature. Tabor, Galantucci, & Richardson (2004) showed that participants are slower to read object-modifying reduced relative clauses (RCs) when the RC verb is part-of-speech ambiguous (*tossed*) than when it is umambiguous (*thrown*). This result is problematic for any fully incremental framework because the main-verb interpretation is incompatible with the global context and should thus be ignored by the processor. An unresolved issue with this result, however, is the extent to which such effects occur in natural reading. We test the extent to which such effects occur in natural reading, using eye-tracking data from free reading of newspaper text. Our results support both of the major theories of local coherences, and add ecological validity to the study of local coherences.

1. Background and Motivation

background

Tabor, Galantucci, & Richardson (2004) varied verb ambiguity and relative clause reduction in object-modifying relative clauses, using sentences such as in (1):

(1) a. The coach smiled at the player tossed a frisbee . . . [+amb, +red]
b. The coach smiled at the player who was tossed a frisbee . . . [+amb, -red]
c. The coach smiled at the player thrown a frisbee . . . [-amb, +red]
d. The coach smiled at the player who was thrown a frisbee . . . [-amb, -red]

- found an interaction such that [+amb,+red] was super-additively hard
- problematic result for any fully incremental framework because the main-verb interpretation is incompatible with the global syntactic context

theories

Tabor et al. (2004):

- difficulty arises because the 'locally coherent' parse of 'the player tossed a frisbee' competes with top-down syntactic expectations

Gibson (2006):

- difficulty arises from word part-of-speech (POS) disambiguation
- difficulty inversely proportional to context-independent probability of tag given word multiplied by the smoothed syntactic expectations

question

- both theories make broad claims about the parser from one example involving a very rare construction, a reduced relative of a passivized ditransitive
- how often do such effects arise in the reading of natural text?

plan

- build a model of the reading times in a corpus of naturalistic eye-tracking data
- test factors representing each of the two major theories

References:

Baayen, R.H., Davidson, D.J., & Bates, D.M. (In Press). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language.*

Charniak, E. (2000). A maximum-entropy-inspired parser. In *Proceedings of the 1st Conference of the North American Chapter of the Association for Computational Linguistics*, (pp. 132–139), Seattle, WA.

Demberg, V., & Keller, F. (To Appear) *Cognition*.

Gibson, E. (2006). The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language*, 54, 363–388.

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, 45, 153–168.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50, 355–370.

2. Corpus, Model, and Control Factors

corpus

- used the Dundee corpus (Kennedy & Pynte, 2005) of eye-movement data from 10 participants reading 51,000 words of *The Independent* (a British newspaper)

- to get syntactic category information, we parsed the corpus with the Charniak parser (Charniak, 2000)

model

each of our factors was tested in a linear mixed-effect model (Baayen, Davidson, & Bates, In Press) of the **first pass times** on each word, with 11 fixed effect control factors and participant as a random effect, as in Demberg & Keller (To Appear)

control factors

Demberg & Keller's control factors included:

linguistic properties:

- word length and frequency
- bigram probability
- position in the sentence
- lexicalized and unlexicalized syntactic surprisal [= -log P(w_i | *context*)]

eye movement properties:

- landing position
- launch distance
- whether the previous word was fixated

3. Factor One: $P(t_i|w_i)$

motivation

Gibson (2006) predicts a word *w* to be read slower with a particular POS tag *t* as the context-independent probability $P(t_i|w_i)$ decreases

estimation

- for each word-tag pair in the Dundee corpus, we estimated $P(t_i|w_i)$ from a Charniak-parsed version of the British National Corpus (BNC)
- two versions of this factor:
  - P_m: maximum likelihood estimate (MLE)
  - P_s: MLE smoothed with a Bayesian prior, which is the type-averaged $P(t_i|w)$ [see box]

Bayesian prior
set to be:

$$P_{pr}(t_i) = \frac{\sum_w P_m(t_i|w)}{\|w\|}$$

where *w* is a word *type*.

smoothing

$$P_s(t_i|w_i) = \frac{c(t_i, w_i) + \beta\left[P_{pr}(t_i)\right]}{c(w_i) + \beta}$$

where β is set to minimize Dundee corpus perplexity

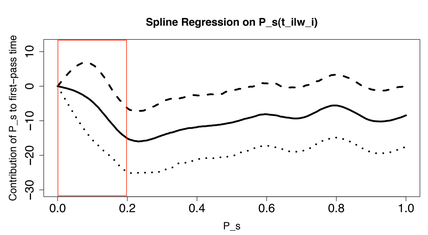
results

tested factor on all words in Dundee corpus that were found in the BNC

	coefficient estimates	
MLE (P_m):	-0.8 [t(199394) = -0.8; $p > .05$]	no effect (MLE too noisy?)
Smoothed (P_s):	-2.3 [t(199394) = -2.2; $p < .05$]	small effect

further analysis

- performed a natural spline regression on $P_s(t_i|w_i)$ with 11 equally spaced knots (see graph; confidence intervals are bootstrapped)
- spline regression suggests that the significant effect is being driven by difference in the 0.0-0.2 range of probability



4. Factor Two: $P(t_i|w_{i-1},w_i)$

motivation

- Tabor et al. (2004) suggest that difficulty can arise when a locally coherent parse conflicts with a top-down parse
- we investigate a subset of these cases by using $P(t_i|w_{i-1},w_i)$ as a factor
- difficulty predicted when $P(t_i|w_{i-1},w_i)$ is low
  - *e.g.*, P(VBN|'player tossed') is low

estimation

- for each word-word-tag triplet in the Dundee corpus, we estimated $P(t_i|w_i)$ from the Charniak-parsed BNC
- two versions: P_m = MLE P_s = MLE smoothed as before

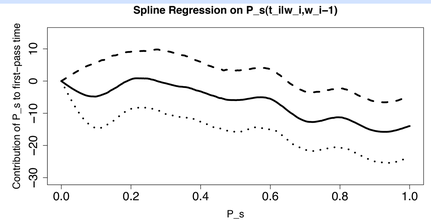
results

tested factor on all words for which the bigram w_{i-1},w_i occurred in the BNC

	coefficient estimates	
MLE (P_M):	-2.9 [t(171556) = -2.28; $p < .05$]	small effect (MLE too noisy?)
Smoothed (P_s):	-17.6 [t(171556) = -16.7; $p < .0001$]	large effect

further analysis

- performed a natural spline regression on $P(t_i|w_{i-1},w_i)$ as before
- spline regression shows evidence of an effect for all values of the factor



5. Caveat

correlations

our factors were somewhat correlated with control factors:

	word frequency	bigram probability
$P_s(t_i|w_i)$:	r = .21	r = .19
$P_s(t_i|w_{i-1},w_i)$:	r = .28	r = .40

- because these control factors were explicitly included in the model, we have some confidence that our effect is 'above and beyond' them
- nevertheless, we plan to address this in the future by:
  - residualizing the covariates
  - constructing more estimates of these and other measures of conflict between top-down and bottom-up probabilities

6. Conclusion

summary

- evidence for both Tabor et al. and Gibson's theories of local coherence effects
- adds ecological validity to the study of local coherences

next step

test a larger subset of cases predicted by Tabor et al. to be difficult by using a modified parser to predict conflict between top-down and bottom-up parses