

## Reconciling Centering-driven and Coherence-driven Accounts of Pronoun Interpretation

Two classic theories of pronoun interpretation have each sought to specify the relationship between pronoun use and discourse coherence, but make seemingly irreconcilable claims. According to Hobbs (1979, 1990), pronoun interpretation is not governed by an independent mechanism, but instead comes about as a by-product of utilizing world knowledge during the inferential establishment of discourse COHERENCE RELATIONS. Factors pertaining to the grammatical form and information structure of utterances do not come into play. According to Centering Theory (Grosz et al. 1986/1995, *inter alia*), on the other hand, pronoun interpretation is predominantly determined by information structural relationships within and between utterances (e.g., topic transitions) and the grammatical roles occupied by potential referents. Factors pertaining to world knowledge and the establishment of informational coherence relations do not come into play.

Recent work in psycholinguistics has supported the Hobbsian view, showing that the semantic biases associated with coherence relations can have a significant impact on pronoun interpretation biases. For instance, Rohde et al. (2006, 2007) showed that while the distribution of pronoun interpretations in passages with transfer-of-possession contexts (1a) is relatively 50/50 between the subject (Source) and non-subject (Goal) referents (confirming Stevenson et al., 1994), they are far from 50/50 when the completions are broken down by coherence relations, with some relations (EXPLANATION, ELABORATION) exhibiting a strong Source bias, and others (OCCASION, RESULT) exhibiting a strong Goal bias.

- (1) a. John passed the comic to Bill. He \_\_\_\_\_  
b. John passed the comic to Bill. \_\_\_\_\_

Left unexplained by the Hobbsian view, however, is evidence that Stevenson et al. (1994) found for an overlaid subject bias, whereby there were more non-subject Goal continuations in the no-pronoun prompt condition (1b) than in (1a). Furthermore, a significant asymmetry emerged in the participants' choices of referential form in (1b), being heavily biased towards a pronoun for references to the subject and a proper name for references to the non-subject.

The foregoing findings prompted Kehler et al. (2008) to propose a Bayesian formulation of pronoun interpretation:

$$(2) \quad P(\textit{referent} \mid \textit{pronoun}) = \frac{P(\textit{pronoun} \mid \textit{referent})P(\textit{referent})}{P(\textit{pronoun})}$$

Equation (2) suggests a reconciliation of the roles of Centering and Hobbsian coherence establishment in determining pronoun interpretation biases. A central tenet of the original Grosz et al. formulation of Centering is that speakers have a *production* bias toward pronominalizing the current topic (i.e., the  $C_b$ ; the previous subject in most examples cited in the literature). This bias is represented by the likelihood  $P(\textit{pronoun} \mid \textit{referent})$ ; existing evidence is consistent with this bias being driven primarily by grammatical or information-structural factors (see also Fukumura and van Gompel, 2010). The semantic biases towards the next-mentioned referent contributed by (Hobbsian) coherence-driven expectations, on the other hand, are represented by the prior  $P(\textit{referent})$ ; the foregoing evidence is consistent with these being largely independent of choice of referential form.

We examine a set of predictions of the analysis using implicit-causality (IC) verbs, which are known to be associated with strong biases regarding which referent will be mentioned next when a causal relation is prompted. For instance, verbs like *amaze* and *scold* are biased toward subsequent reference to the subject and object in (3) respectively (Au 1986, Stevenson et al. 1994, inter alia).

(3) Amanda amazed/scolded Brittany because she \_\_\_\_\_

Kehler et al. (2008) also showed that IC verbs carry an additional bias. In similar passages with full-stop prompts (see 4c below), participants continue IC contexts by describing a cause of the preceding eventuality (i.e., Explanation coherence relations) much more often (60%) than for non-IC verbs (24%). With this in mind, we designed contexts with subject-biased IC verbs as in (4):

- (4) a. Amanda amazed Brittany. She \_\_\_\_\_  
b. Brittany was amazed by Amanda. She \_\_\_\_\_  
c. Amanda amazed Brittany. \_\_\_\_\_  
d. Brittany was amazed by Amanda. \_\_\_\_\_

The coherence-driven and Centering pronoun biases converge in cases like (4a), both pointing to the causally-implicated referent, Amanda, as the preferred interpretation of *she*. These biases are opposed in (4b), however, where the IC verb has been passivized. If the structural voice alternation – which preserves the proposition denoted by the sentence, but establishes the causally non-implicated referent as the topic – is irrelevant to both pronoun biases and the likelihood of generating Explanation continuations (as it should be on Hobbs’ account), we should see the same distributions for (4b) as (4a). Alternatively, if comprehenders integrate Centering biases along with coherence-driven ones during interpretation, then two predictions emerge: (i) we should see fewer continuations in (4b) than (4a) in which the pronoun refers to the causally-implicated referent, and (ii) the added bias against this referent should mean fewer Explanation continuations in (4b) than (4a). The full-stop prompts in (4c-d) were included to ensure that the previously reported subject bias is replicable in these contexts.

As predicted, regardless of voice, participants (N=42) were significantly more likely to write a continuation that referred to the subject of the previous sentence in the pronoun-prompt condition than the no-pronoun condition, replicating previous studies (active: 76.8%/58.8%; passive: 58.0%/22.2%). Further, the two main predictions for the pronoun-prompt data were confirmed: (i) the bias toward the IC-favored referent was significantly weaker in the passive pronoun condition (4b; 42.0%) than the active one (4a; 76.8%), and (ii) there were significantly fewer Explanation continuations in the passive pronoun condition (4b; 52.4%) than in the active one (4a; 74.6%). As such, even in contexts with strong semantic biases, the mere occurrence of a (fully-ambiguous) pronoun not only shifts interpretation biases toward the subject referent, but also influences comprehenders’ expectations about how the discourse will be coherently continued. These data therefore support an account in which pronoun interpretation and its relationship to discourse coherence are co-determined by production-oriented Centering biases concerning choice of referential form and coherence-driven expectations about what entities will be subsequently mentioned.

## References

- Au, Terry K. 1986. A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language*, 25:104–122.
- Fukumura, Kumiko and Roger P. G. van Gompel. 2010. Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language*, 62:52–66.
- Grosz, Barbara J., Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.
- Hobbs, Jerry R. 1979. Coherence and coreference. *Cognitive Science*, 3:67–90.
- Hobbs, Jerry R. 1990. *Literature and Cognition*. CSLI Lecture Notes 21, Stanford, CA.
- Kehler, Andrew, Laura Kertz, Hannah Rohde, and Jeffrey L. Elman. 2008. Coherence and coreference revisited. *Journal of Semantics (Special Issue on Processing Meaning)*, 25(1):1–44.
- Rohde, Hannah, Andrew Kehler, and Jeffrey L. Elman. 2006. Event structure and discourse coherence biases in pronoun interpretation. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, pages 697–702, Vancouver, Canada.
- Rohde, Hannah, Andrew Kehler, and Jeffrey L. Elman. 2007. Pronoun interpretation as a side effect of discourse coherence. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society*, pages 617–622, Memphis, TN.
- Stevenson, Rosemary J., Rosalind A. Crawley, and David Kleinman. 1994. Thematic roles, focus, and the representation of events. *Language and Cognitive Processes*, 9:519–548.