

Evaluating a Coherence-Based Model of Pronoun Interpretation

Laura Kertz, Andrew Kehler, Jeffrey L. Elman

Abstract

We describe two pronoun interpretation experiments in which a Coherence Hypothesis is tested against preference-based accounts. The Coherence Hypothesis holds that apparent preferences in antecedent selection are actually byproducts of the inferencing processes used to establish different types of coherence. In Experiment 1 we show that preferences can be systematically disrupted through the manipulation of coherence and that when the relevant factors are balanced, preferences disappear. In Experiment 2 we show that the coherence effect is not disrupted by voice alternations (active/passive), providing evidence for a strong semantic model of coherence-driven interpretation. We speculate on the adequacy of this strong semantic model and propose additional online experiments to examine the interaction between propositional content and information structure influences on pronoun interpretation.

1. Introduction

The pronoun interpretation literature over the last three decades has followed two main lines of investigation. One approach casts pronoun interpretation as a matching process guided by ‘heuristics’, that is, broadly-applicable strategies for matching pronouns to their antecedents. Of particular relevance to our purposes is a body of work developed in the 1990’s that examined two competing preferences: the parallel function preference (Smyth 1994, Chambers & Smyth 1998, *inter alia*) and the subject assignment preference (Crawley & Stevenson 1990, Crawley et al. 1990, *inter alia*). Centering Theory (Grosz et al. 1995 [1986], Brennan, Friedman, and Pollard 1987) incorporates neither of these preferences directly, but shares the property that pronoun interpretation is driven in large part by relationships based on grammatical roles.

In this paper, we describe two psycholinguistic experiments that support a model in which the interpretation heuristics posited in the literature are epiphenomena of processes that hearers use to establish discourse coherence. In Experiment 1, we demonstrate (following Kehler 2002) that such preferences -- in particular, the parallel function preference and the subject assignment preference -- are not reliable indicators of pronoun interpretation. In Experiment 2, we test the interaction of coherence with referent accessibility using a voice (active/passive) manipulation. We find that in an off-line task, coherence trumps accessibility. We conclude with a discussion of the possible limitations of these findings, and suggest further research using online measures.

2. Preferences and Coherence

Heuristics-based models of pronoun interpretation predict that morpho-syntactic cues can generally be used to identify a single preferred antecedent. These heuristics, however, are often in conflict. For example, the subject assignment preference identifies the preferred antecedent as the subject of the preceding clause, a condition which holds in (1a) and (1c), but not in (1b) and (1d). The parallel function preference, on the other hand, predicts that an antecedent will appear in the same argument position as the pronoun, a claim supported by (1a) and (1b), but not in (1c) or (1d). A modified version of the parallel function preference that requires full syntactic parallelism between the clauses (Smyth 1994) suggests that the presence/absence of a modifier, as in (1a-d), may shift these preferences.

(1)	Samuel threatened Justin with a knife, and		<i>reference</i>	<i>coherence</i>
a.	he blindfolded Erin (with a scarf).	(=Samuel)	parallel	parallel
b.	Erin blindfolded him (with a scarf).	(=Justin)	parallel	parallel
c.	Erin stopped him (with pepper spray).	(=Samuel)	non-parallel	cause/effect
d.	he alerted security (with a shout).	(=Justin)	non-parallel	cause/effect

The preference accounts stand in contrast to a coherence-driven model, as originally proposed by Hobbs (1979), which suggests that pronoun interpretation is not driven by morpho-syntactic cues but rather is the by-product of larger inferencing processes that support the establishment of discourse coherence. Kehler (2002) extends Hobbs's proposal, arguing that interpretation 'preferences' are actually epiphenomena of the manner in which different types of coherence are established. Preliminary support for this model was reported by Wolf et al. (2004), which we extend in several ways here. Specifically, we test Kehler's (2002) predictions in behavioral experiments, focusing on two coherence types: parallel coherence, which obtains between clauses with similar propositional content, and cause/effect coherence, which obtains between clauses denoting events that are causally linked.

3. Experiment 1

Kehler (2002) suggests that the conflicting preferences reported in the psycholinguistics literature often result from a failure to control for coherence across stimuli. To address this, we constructed stimulus sets which exhibit parallel coherence in half the cases and cause/effect coherence in the other half. As in Wolf et al, the cause/effect cases incorporated a semantic bias toward the non-parallel referent; the parallel cases incorporated no bias. Coherence was assessed during a prior norming study. Note that we used only *and* as a connective to guard against potential 'focusing' effects of connectives like *and similarly* or *and so* as claimed by Stevenson et al. (2000). We also varied the choice of verb class to guard against the possibility of a thematic role bias (Stevenson et al. 1994, Stevenson et al. 2000).¹

To test the parallel function preference, we also varied pronoun position. Recall that while the subject assignment strategy predicts an across-the-board preference for subject

¹ Verbs in the introductory clause were drawn from one of four classes: physical action (threaten, punch, check, injure); social action (deceive, taunt, harass, salute); mental state (demoralize, admire, irritate, intimidate); and speaking verbs (bait, defend, suggest, ridicule). See Rohde et al (2006) for studies examining the interaction of thematic roles, event structure, and coherence.

antecedents, the parallel function preference predicts a subject preference for subject pronouns, and an object preference for object pronouns (a main effect of pronoun position). The modified parallel function preference suggests that this effect may be mitigated when the two clause structures are not perfectly parallel (an interaction between pronoun position and syntactic structure). To test this, we varied syntactic structure, creating a perfectly parallel condition in which an adverbial modifier in clause 2 matches the modifier in clause 1 (i.e., the parentheticals in 1a-d are included), and a partially parallel condition in which no modifier appears in clause 2 (i.e., the parentheticals in 1a-d are not included). The Coherence Hypothesis predicts parallel coreference in the parallel coherence condition, and non-parallel reference in the cause/effect coherence condition (i.e. an interaction between coherence and pronoun position).

3.1 Method

- Participants were 32 undergraduates who received extra credit for participation. All were self-reported mono-lingual native speakers of English.
- In a 2x2x2 design, we constructed 16 stimulus sets, each with 8 variants, for a total of 128 stimuli, as described above. Verb class and modifier type were also balanced across stimuli. A sample stimulus set is presented above in (1).
- In a repeated measure design, participants were tested on two instances of each stimulus type, but no participant saw more than one variant from the same set. The 16 experimental stimuli were interleaved with 24 distracters (16 stimuli from experiment 2; and 8 fillers with unambiguous pronouns), and the replications were block randomized.
- Participants were presented with a paper and pencil task in which the stimulus was followed by a question requiring that the pronoun be interpreted, as in (2). For answers matching the subject of the preceding clause, a score of 1 was assessed. For answers matching the object of the preceding clause, a score of 0 was assessed.

(2) Samuel threatened Justin with a knife, and he blindfolded Erin with a scarf.
Who blindfolded Erin?

3.2 Results

The data showed strong support for the Coherence Hypothesis, confirming the predicted interaction between pronoun position and coherence frame. The data were not consistent with any of the other hypotheses. Gross percentages across stimuli show that the ‘preferences’ operate at roughly chance levels when the relevant factors (including coherence) have been balanced.

Table 1 Experiment 1

<i>Hypothesis</i>	<i>Condition</i>	<i>Antecedent</i>		<i>n</i>
		<i>Subj</i>	<i>Obj</i>	
Subject Assignment	All pronouns	.52	.48	512
Parallel Function	Subject pronouns	.51	.49	256
	Object pronouns	.52	.48	256
Modified Parallel Function	Subject pronoun: fully parallel structure	.52	.48	128
	Object pronoun: fully parallel structure	.50	.50	128
Coherence Hypothesis	Subject pronoun: Parallel coherence	.98	.02	128
	Subject pronoun: Result coherence	.05	.95	128
	Object pronoun: Parallel coherence	.10	.90	128
	Object pronoun: Result coherence	.94	.06	128

Statistical analysis of these data confirms that the interaction between coherence frame and pronoun position, predicted by the Coherence Hypothesis, is significant [$F(1,31) = 1379.23$, $p < .0001$; $F(1,15) = 2016.158$, $p < .0001$]. A second, smaller effect, which we did not predict, was found for coherence alone [$F(1,31) = 4.429$, $p = .044$; $F(1,15) = 7.105$, $p = .018$].

4. Experiment 2

The results of Experiment 1 support the Coherence Hypothesis. They show that pronoun interpretation does not depend exclusively on pronoun position or syntactic structure, both of which were balanced across stimuli. Nor was the preferred antecedent signaled by discourse connectives like *and similarly* or *and so* (as suggested by Stevenson et al. 2000). Rather, pronoun interpretation depended primarily on the semantic content of the propositions and whether the relationship between them is based on similarity or causality.

While the results of Experiment 1 appear to support the type of ‘strong’ coherence-based model we have described, they do not rule out the possibility of a ‘weaker’ coherence model, in which coherence interacts with other factors such as information structure. To differentiate these possibilities, Experiment 2 tested participants on passages in which the voice of the context sentence was manipulated. Because the passive marks the surface subject as topical (and likewise reduces the topicality of the logical subject), we might expect support for pronominal reference to the logical subject via coherence to be compromised in such constructions.

Thus, the strong semantic model predicts the replication of the Experiment 1 results in both the active and the passive conditions: parallel co-reference with parallel coherence; non-parallel co-reference with cause/effect coherence. (Note, importantly, that parallelism now refers to thematic role, not argument position, i.e. semantically parallel, not syntactically). Thus, when we exclude pronoun position as a factor (using only object pronouns), the choice of antecedent (subject/non-subject) will depend on an interaction between coherence and voice. The weak model also predicts an interaction between coherence and voice, but it predicts a different pattern, one in which the clean results seen in Experiment 1 are

replicated in the active voice, but where the interaction is weakened in the passive, with voice playing a greater role.

4.1 *Method*

- Participants were the same as in Experiment 1.
- In a 2x2 design, we adapted 16 stimulus sets from Wolf et al. (2004), constructing 4 variants for each set, for a total of 64 stimuli. Stimuli contained all object pronouns, but varied in coherence frame (parallel vs. cause/effect; based on verb semantics) and voice (active/passive). A sample stimulus set is presented in (3).
- In a repeated measure design, participants were tested on four instances of each stimulus type. All remaining aspects of the design, including task, were identical to Experiment 1.

- (3) a. James complimented Craig, and Fiona congratulated him after the match. parallel coherence
active voice
- b. James defeated Craig, and Fiona congratulated him after the match. cause/effect coherence
active voice
- c. James was complimented by Craig, and Fiona congratulated him after the match. parallel coherence
passive voice
- d. James was defeated by Craig, and Fiona congratulated him after the match. cause/effect coherence
passive voice

4.2 *Results*

The data show a clear interaction between coherence and voice, predicted by both models. The pattern in the data, however, is symmetrical for active and passive, supporting the strong coherence model. Passivization does not appear to weaken the interaction between coherence and voice: we observe parallel reference in the parallel coherence frame, and non-parallel reference in the cause/effect frame. (The co-reference pattern reverses, however; in the passive, the patient role is realized in subject position.)

Table 2 Experiment 2

<i>Condition</i>		<i>Antecedent</i>		
		<i>Subj</i>	<i>Non-Subj</i>	<i>n</i>
parallel coherence:	active voice	.22	.78	128
cause/effect coherence:	active voice	.84	.16	128
parallel coherence:	passive voice	.77	.23	128
cause/effect coherence:	passive voice	.23	.77	128

Statistical analysis of these data confirms that the interaction between coherence frame and voice is significant [$F(1,31) = 278.52, p < .0001$; $F(1,15) = 118.64, p < .0001$]. In the item analysis, an additional interaction was detected among item, voice, and coherence [$F(1,15) = 2.712, p = .0005$], indicating some degree of variability in the stimulus set, which we are currently investigating.

5. Discussion

The results of Experiment 1 confirm the prediction in Kehler (2002) that pronoun interpretation preferences can best be understood as epiphenomena of coherence-based inferencing processes. Parallel coherence results in parallel co-reference, whereas causal inference determines co-reference in cause/effect coherence frames. The Coherence Hypothesis makes the correct predictions in cases in which the preference accounts conflict, and also predicts the subject-pronoun-to-object-antecedent pattern in the cause-effect passages that contradict both preference hypotheses. (See Table 3.) Further, the results of Experiment 2 suggest that this strong coherence-driven model of pronoun interpretation is resilient to the passive/active voice distinction.

Table 3 Predicted Resolution Patterns (Pronoun to Antecedent)

<i>Hypothesis</i>	<i>subject to subject</i>	<i>subject to object</i>	<i>object to object</i>	<i>object to subject</i>
Subject Preference:	yes	no	no	yes
Parallel Preferences:	yes	no	yes	no
Coherence:	yes	yes	yes	yes

Whereas the effects appear to be robust in parallel and cause/effect coherence frames, at this point we cannot be certain that it will generalize to other coherence types. It may be the case that parallel and cause/effect coherence relations rely more heavily on semantic content than do other forms of coherence, for example, narrative coherence (e.g., Hobbs's Occasion relation).

A variety of effects associated with the subject preference, topic preference, and first mention effects (among others reported in the literature), appear to be attention-based (cf. various forms of Centering Theory). Kehler (2002) suggests that attentional effects may be most prominent in Occasion frames, in which a series of events connected in time and/or space are described. For example, in a stimulus like (4), the most likely coherence frame in the final sentence is Occasion.

- (4) Betty and Sally had been pestering the firm for months to buy a new photocopier. Duncan was hoping that he would be able to use it too. Betty demonstrated the new machine to Sally and Duncan asked her about it. (Crawley et al. 1990)

While it is not impossible to draw either a causal or similarity-based connection between the demonstration and the questioning, such inferences are much weaker than the sort drawn in the parallel and cause/effect variants of example (1). It is possible that Occasion is intertwined with notions such as topichood and salience to a greater degree than are the other relations. Experiment 2 did not show an interplay between attention and coherence, but if such an effect can be demonstrated, this would support a richer model of coherence recognition and pronoun interpretation than previously described (for example, in Hobbs 1979).

6. Future Work

While our off-line experiments showed that participants identify the referents that are supported by coherence establishment, they provide no measure of the difficulty with which those identifications were made. As such, experiments currently in preparation will use manipulations similar to those described here, but embedded within a self-paced reading task. We hope that this online measure will allow a fuller view of the processes involved in interpretation than the offline disambiguation task. In particular, we will examine cases like (1a) and (1d), repeated here as (5) and (6), in which the semantic information supporting coherence appears after the pronoun. We are particularly interested in whether any biases introduced by the subject assignment and/or parallel function preferences cause examples like (5) to be easier to process than those like (6), in which both heuristics point to the semantically dispreferred referent.

- (5) Samuel threatened Justin with a knife, and he blindfolded Erin (with a scarf).
(6) Samuel threatened Justin with a knife, and he alerted security (with a shout).

We will also look for reading time delays in cases like (3d) from Experiment 2, repeated here as (7), for which we predicted a clash between information structure (supporting the subject antecedent) and coherence (supporting a non-subject antecedent).

- (7) James was defeated by Craig, and Fiona congratulated him after the match.

Finally, the norming phase of Experiment 1 turned up a variety of interesting data relevant to the question of coherence recognition. These data suggest various ways that aspects of linguistic form and structure might affect the coherence relations that hearers identify, beyond the effects of semantics proper. We are currently developing experiments that control for these effects and initiating an investigation into the processes that drive coherence recognition.

References

- Brennan, S. E., Friedman, M. W., & Pollard, C. J. (1987). A Centering Approach to Pronouns. Proceedings of the 25th Meeting of the Association for Computational Linguistics, 155-162.

- Chambers, C. C., & Smyth, R. (1998). Structural Parallelism and Discourse Coherence: A Test of Centering Theory. *Journal of Memory and Language*, 39, 593-608.
- Crawley, R. A., & Stevenson, R. J. (1990). Reference in Single Sentences and in Texts. *Journal of Psycholinguistic Research*, 19, 191-210.
- Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The Use of Heuristic Strategies in the Interpretation of Pronouns, *Journal of Psycholinguistic Research* 19, 245-264.
- Grosz, Barbara J., Joshi, A. K., & Weinstein, S. (1995 [1986]). Centering: A Framework for Modelling the Local Coherence of Discourse, *Computational Linguistics* 21:2, 203-225.
- Hobbs, J. R. (1979). Coherence and Coreference, *Cognitive Science* 3, 67-90.
- Kehler, A. (2002). *Coherence, Reference, and the Theory of Grammar*. Stanford, CA: CSLI Publications.
- Rohde, H., Kehler, A., and Elman, J. L. (2006). Event Structure and Discourse Coherence Biases in Pronoun Interpretation. To appear in the Proceedings of the 28th Annual Conference of the Cognitive Science Society, Vancouver, Canada.
- Smyth, R. (1994). Grammatical Determinants of Ambiguous Pronoun Resolution. *Journal of Psycholinguistic Research*, 23, 197-229.
- Stevenson, R. J., Crawley R., & Kleinman D. (1994). Thematic Roles, Focusing and the Representation of Events. *Language and Cognitive Processes* 9, 519-548.
- Stevenson, R. J., Knott, A., Oberlander, J., & McDonald, S. (2000). Interpreting Pronouns and Connectives: Interactions among Focusing, Thematic Roles, and Coherence Relations, *Language and Cognitive Processes* 15:3, 225-262.
- Wolf, F., Gibson, E. & Desmet, T. (2004). Discourse Coherence and Pronoun Resolution. *Language and Cognitive Processes*, 19:6, 665-675.

Laura Kertz
Department of Linguistics, 9500 Gilman Drive
La Jolla, CA 92093-0108 USA
kertz [at] ling.ucsd.edu

Andrew Kehler
Department of Linguistics, 9500 Gilman Drive
La Jolla, CA 92093-0108 USA
kebler [at] ling.ucsd.edu

Jeffrey L. Elman
Center for Research in Language, 9500 Gilman Drive
La Jolla, CA 92093-0515 USA
elman [at] crl.ucsd.edu