Pronoun Interpretation as a Side Effect of Discourse Coherence

Hannah Rohde (hannah@ling.ucsd.edu)
Andrew Kehler (kehler@ling.ucsd.edu)
Department of Linguistics; University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0108, USA

Jeffrey L. Elman (jelman@ucsd.edu)
Department of Cognitive Science; University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0515, USA

Abstract

Recent story completion studies (Arnold 2001, Rohde et al. 2006) show that passages exhibiting different coherence relations yield different patterns of pronoun interpretation. These results predict that a shift in the distribution of coherence relations in participant responses ought to induce a shift in the distribution of pronoun interpretations. Experiment 1 manipulates the coherence distribution by varying the direct object in Rohde et al.'s original stimuli. We find that the likelihood of a pronoun referring to a specific referent varies by relation and that the conditional probabilities (*p(referent|coherence)*) remain consistent across conditions. In Experiment 2, we vary only the instructions, having participants write continuations that answer either the question 'Why' (Explanation relation) or 'What happened next?' (Occasion). As predicted, both pronoun interpretation and the coherence distribution differ significantly by instruction type, and the pattern of pronoun interpretation corresponds directly to the distribution of coherence relations. Models of pronoun interpretation that ignore discourse coherence relationships cannot account for results like these.

Keywords: discourse processing, pronoun interpretation

Interpreting Ambiguous Pronouns

Whereas previous work on pronoun interpretation has appealed to surface-level cues like subjecthood, first-mention, recency, and parallelism, recent work suggests that interpretation should be analyzed in part as a byproduct of deeper discourse-level comprehension processes. Rohde, Kehler, & Elman (2006) mimicked the design of a story continuation study by Stevenson, Crawley, & Kleinman (1994) in order to show that pronoun interpretation differs in transfer-of-possession passages that vary by verbal aspect.

(1) John$_{SOURCE}$ handed a book to Bob$_{GOAL}$. He _____.

(2) John$_{SOURCE}$ was handing a book to Bob$_{GOAL}$. He _____.

The context sentences in (1) and (2) contain two possible referents for the pronoun, one that appears in subject position and fills the Source thematic role, and one that appears as the object of a prepositional phrase and fills the Goal thematic role. Stevenson et al.'s work, based on stimuli like (1), showed that participants wrote just as many continuations that corresponded to a Goal interpretation for the pronoun as to a Source interpretation. Two explanations were considered: a thematic-role preference (favoring Goals over Sources) and an event-structure hypothesis (a bias to focus on the end state, where the Goal is assumed to be more salient than the Source in transfer passages). Rohde et al.'s aspect manipulation distinguished between these two explanations because the thematic roles remain the same in (1) and (2), but the perfective verb in (1) describes a completed event, which is compatible with end state focus, while the imperfective verb in (2) describes an ongoing event that lacks an end state. They found that perfective sentences yielded more Goal continuations than imperfective sentences, thereby supporting Stevenson et al.'s event-structure hypothesis.

Rohde et al.'s work further supports a model which incorporates deeper discourse-level factors. Specifically, they found that the influence of event structure was observed only in passages in which certain relationships could be inferred to hold between the two clauses (see Arnold 2001 for similar results). These results predict that a shift in the distribution of such relationships (henceforth 'coherence relations') ought to induce a shift in the distribution of pronoun interpretations. If people mainly process ambiguous pronouns using surface-level biases, one might expect their continuations to expose these biases since they are free to write continuations that are consistent with the preferred interpretation. If, on the other hand, participants are sensitive to coherence-driven factors that allow them to generate predictions about where a discourse is going, they may use the pronoun differently depending on the context.

In this paper, we present two experiments that manipulate the coherence distribution in story continuations by making only minimal changes to Rohde et al.'s original stimuli. Our results show that the pattern of pronoun interpretation corresponds directly to the distribution of coherence relations. No model of interpretation that relies entirely on surface-level cues can account for these results. From the story continuations, we derive estimates for the likelihood that a pronoun refers to a specific referent given a coherence relation and discuss how conditional probabilities (*p(referent|coherence)*) can lay the groundwork for the an expectation-driven model of pronoun interpretation.

The Role of Coherence

Coherence relations (Hobbs 1990, Kehler 2002) describe the relationship that is inferred to hold between adjacent clauses in a discourse. For instance, consider the Occasion relation:

> OCCASION: Infer a change of state for a system of entities in S_2, establishing the initial state for this system from the final state of S_1. (adapted from Hobbs 1990)
>
> (3) Miriam sent a fruitcake to Rachel. She <u>ate it voraciously</u>.   [Occasion (*Goal*)]

Establishing Occasion relations requires that a state of affairs be inferred as a point of connection between the two eventualities expressed by two clauses, i.e., that the initial state of the second eventuality be identified with the final state of the first. Since receiving something is a precondition to eating it, such inference supports interpreting the pronoun in (3) to refer to the Goal Rachel. Rohde et al. report that perfective context sentences were followed most frequently by continuations that resulted in an Occasion relation. The fact that end state salience is a natural consequence of the definition of an Occasion relation explains the large number of Goal continuations observed in those contexts.

On the other hand, Rohde et al. report that context sentences with imperfective verbs were followed more frequently by Explanation and Elaboration relations.

> EXPLANATION: Infer *P* from the assertion of sentence S_1 and *Q* from the assertion of sentence S_2, where $Q{\rightarrow}P$.
>
> (4) Miriam was sending a fruitcake to Rachel. She <u>heard it was a nice thing to do.</u>   [Exp (*Source*)]
>
> ELABORATION: Infer the same proposition P from the assertions of S_1 and S_2.
>
> (5) Miriam was sending a fruitcake to Rachel. She <u>packed it up and went to UPS.</u>   [Elab (*Source*)]

Establishing an Explanation relation requires that the hearer draw a chain of causal inferences that connect the two eventualities that are being related, where the second clause describes the cause of the eventuality in the first. Such inference supports interpreting the pronoun in (4) to refer to the Source Miriam. (Another relation, Result, is similar to Explanation but the cause precedes the effect.) The sentences in Elaboration relations, on the other hand, describe the same event (typically where the second sentence provides more detail than the first), supporting a Source interpretation in (5). Rohde et al. report a Source bias for both Explanation and Elaboration relations.

Given these biases, we predict that increasing the number of Explanation or Elaboration relations will result in a corresponding increase in the number of Source continuations. Similarly, an increase in the number of elicited Occasion or Result[1] relations ought to increase the number of Goal continuations. These predictions follow Kehler's suggestion that pronoun interpretation may be in part a side effect of the process of establishing coherence.

---

[1] Although the numbers were small, Result relations were also reported to exhibit a Goal bias in Rohde et al.'s experiment (2006).

Experiment 1: Object Type

In an attempt to shift the distribution of coherence relations in participant responses, we modified Rohde et al.'s stimuli to include different types of objects of transfer, anticipating that certain objects might change how participants chose to continue the passage. We paired each normal object from Rohde et al.'s original stimuli with a bizarre, surprising, gross, or taboo object—an irrelevant manipulation for any surface-cue theory. As such, the structure of the transfer-of-possession event remains unchanged and the grammatical positions of the Source and Goal referents remain constant. The only difference is the normality of the object of transfer.

(6) NORMAL TRANSFER OBJECT
    John gave a **book** to Bob. He _____.

(7) ABNORMAL TRANSFER OBJECT
    John gave a **bloody meat cleaver** to Bob. He ___.

We predicted that abnormal objects would elicit more continuations that explained the event in the first sentence. Our hope was that the increase in (Source-biased) Explanations would be at the expense of (Goal-biased) Occasion relations, which were the most frequent relation following mundane objects in Rohde et al. As such, we expected to find an increase in the number of Source interpretations in continuations following abnormal objects.

Participants   Sixty-nine monolingual English-speaking undergraduates at UC San Diego participated in the study for extra credit in Linguistics courses.

Materials   The experimental stimuli, taken directly from Rohde et al., consisted of twenty-one transfer-of-possession context sentences followed by an ambiguous pronoun prompt. Participants saw each transfer verb only once, paired with either a normal or abnormal object of transfer and using perfective or imperfective aspect. We included the same twenty-nine fillers from Rohde et al.'s stimuli. Fillers consisted of context sentences with transitive or intransitive non-transfer verbs in the perfective or imperfective. The transitive verbs varied in active and passive voice. Adverbs, names, or gender-unambiguous pronouns served as prompts.

Task   Participants were instructed to imagine a natural story continuation, writing the first continuation that came to mind and avoiding humor. In this task, participants create a mental model of the event in the context sentence and then write a continuation that reflects their expectations about where the story is going. As such, the task involves both interpretation and production. The pronoun constrains the surface realization of their continuation, but the continuation depends on their expectations about how the discourse will proceed and which individual will be mentioned again.

Evaluation and Analysis   Following Rohde et al., two trained judges assessed the participants' intended pronoun interpretation. Judges were instructed to err on the side of categorizing a pronoun as ambiguous if the pronoun could

be interpreted plausibly as coreferential with either referent.[2] Disagreements were resolved through discussion, following Stevenson, Knott, Oberlander, & McDonald (2000), and ambiguous cases were set aside. Judges identified the coherence relation that held between each context sentence and continuation. Passages were annotated for six relations including the four already mentioned (Occasion, Explanation, Elaboration, and Result) as well as two others: Violated Expectation (another cause-effect relation) and Parallel (which patterns most closely with Elaboration). The coherence relations were defined as in Kehler (2002).

For the analysis, we measured the effects of two within-subjects factors (object type and verbal aspect) on two dependent variables (the two-valued pronoun interpretation and the six-valued coherence relation). Analyses of variance were conducted on the assessed pronoun interpretations with tests for main effects of verbal aspect and object type, as well as an interaction between the two. The distribution of coherence relations was modeled using a mixed-effects multinomial logistic regression (Bates 2006). These models, like any linear regression, attempt to fit a linear model to a set of data by estimating parameter values for each predictor in the model. They allow us to simultaneously model subjects and items as random factors, and they allow us to test for effects on an unordered noncontinuous dependent variable, the coherence relation.[3] Like other regressions, the model is evaluated by its ability to predict the dependent variable; the effects of individual factors are measured with p-values that represent the likelihood of a model generating a non-zero factor coefficient if the true coefficient were zero. We report the coefficient and p-value for each factor in a model fitted to the observed coherence distribution.

Results and Interpretation

As in Rohde et al., the distribution of pronoun interpretation and coherence relations differed following perfective and imperfective context sentences (significant effect of aspect on interpretation: by subject: $F(1,68)=94.459$, $p<0.0001$; by item: $F(1,20)=98.222$, $p<0.0001$; in logistic regressions, aspect was a significant factor for modeling interpretation: 2.001, $p<0.0001$ and modeling coherence: 1.299, $p<0.0001$).

As predicted, the coherence distribution also differed by object (object coefficient in regression: 0.574, $p<0.05$).[4] Figure 1 shows the coherence breakdown of continuations following context sentences with different object types.

---

[2] Our use of judges follows Arnold (2001). Stevenson et al. (1994) had participants circle their intended referents after completing the passages, but they nonetheless found it necessary to rely on judges to verify participants' circling.
[3] ANOVA is inappropriate because our dependent variable has six unordered categorical outcomes. Chi-square can be used to compare distributions, but our repeated-measures design necessarily includes multiple observations from each subject, thereby violating the chi-square assumption of independence.
[4] The traditional, but in this case inappropriately applied, chi-square test also shows the coherence distribution differing by object ($\chi^2 =28.810$, $p<0.0001$) and aspect ($\chi^2=44.109$, $p<0.0001$).
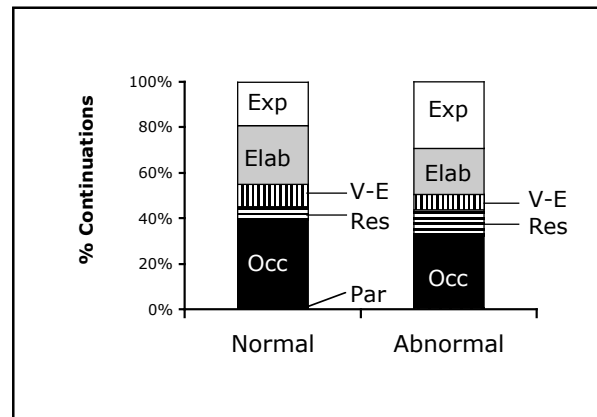

Figure 1: Effect of object type on coherence relations

As predicted, abnormal objects did yield more Explanation relations and fewer Occasion relations than normal objects – but there were other shifts in the coherence distribution as well. The first column of numbers in Table 1 shows the interpretation biases Rohde et al. found for each coherence relation: Elaboration/Explanation/Violated-Expectation are Source-preferring; Occasion/Result are Goal-preferring.[5] Factoring together these biases and the coherence distribution in Figure 1 results in no net increase in the proportion of Source-biased relations: The proportion of Explanation relations increased at the expense of another Source-biased relation, Elaboration, and the proportion of Occasions decreased as the number of Results, another Goal-biased relation, increased. If the biases in this experiment are like those in Rohde et al., which one would expect, then we predict similar proportions of Source and Goal interpretations following normal and abnormal objects.

Table 1: Conditional probabilities collapsed across conditions

Coherence relation	Prob(Source | Relation)	
	Rohde et al. 2006	Exp 1
Elaboration	.98	.99
Explanation	.75	.82
Violated-Expectation	.91	.81
Occasion	.19	.17
Result	.16	.05

As the second column of probabilities in Table 1 shows, the interpretation biases are consistent with Rohde et al., meaning that we predict essentially no difference in the number of Source interpretations. As Figure 2 shows, this is borne out: object type was not a significant factor for modeling pronoun interpretation ($F<1$ by subject and item; object type coefficient in regression: 0.434, $p<0.391$).[6] There were no interactions ($F<1$ by subject and item).

---

[5] Here and in Rohde et. al's experiment, there were too few cases of Parallel relations (n<15) to allow for generalization.
[6] All reported analyses represent a conservative treatment of the data, excluding cases that were judged to be ambiguous by at least one judge (23% of the total for Experiment 1).
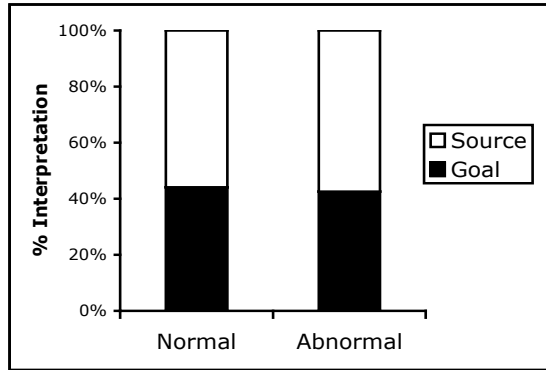
Figure 2: Effect of object type on interpretation

Given these results, we sought an alternative means of creating a shift in the relevant coherence relations. Experiment 2 uses the original stimuli from Rohde et al. unchanged in two conditions. The two conditions differ instead by one sentence in the instructions. This manipulation is again irrelevant to surface-cue theories, but we find that the pattern of interpretation differs in a way that is fully predictable given the distribution of coherence relations and the probabilities summarized in Table 1.

Experiment 2: Instruction Type

Aside from the instructions, Experiment 2 directly mimics Rohde et al.'s original study. Depending which group they were assigned to, participants were instructed to write a continuation that answered either the question "Why" or "What happened next" – the aim being to elicit Explanation and Occasion relations, respectively. Given the interpretation biases from Experiment 1, we predicted that these two types of relations would correspond to very different patterns of pronoun interpretation, making instruction type a good predictor of interpretation.

Participants   Forty-two monolingual English-speaking undergraduates at UC San Diego participated in the study for extra credit in Linguistics courses.

Materials   The twenty-one experimental stimuli and the twenty-nine fillers were taken directly from Rohde et al.

Task   As before, participants wrote continuations for fifty passages. Depending on the instruction condition, they were asked to imagine a natural continuation that answered either the question "Why" or "What happened next".

Evaluation and Analysis   Again, two trained judges assessed the participants' intended pronoun interpretation and the coherence relation that held between the context sentence and the continuation. For the analysis, we were interested in measuring the effects of one between-subjects factor (instruction type) and one within-subjects factor (verbal aspect) on two dependent variables (the two-valued pronoun interpretation and the six-valued coherence relation). For mixed-model repeated measures analyses like this, a logistic regression can again be used to fit the data and identify significant parameters.

Results and Interpretation

The manipulation of instruction type succeeded in altering the distribution of coherence relations. In a logistic regression of the distribution of coherence relations, the between-subject factor of instruction type was significant (coefficient -1.886, p<0.0002).[7] As Figure 3 shows, participants who answered the question 'What happened next' wrote more Occasion relations, Violated Expectations, and Elaborations, while those who answered the question 'Why' wrote mostly Explanations.[8]
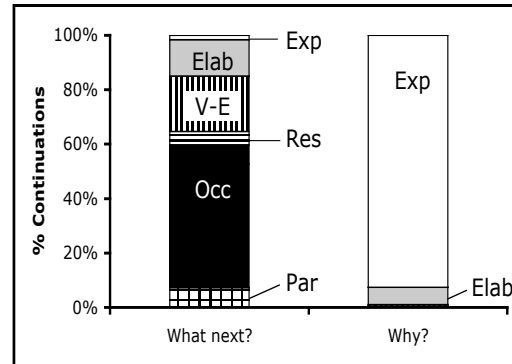

Figure 3: Effect of instruction type on coherence relations

And as predicted, the shift in coherence relations is accompanied by a corresponding shift in the distribution of pronoun interpretations ($F(1,40)=23.302$, p<0.0001 by subject; $F(1,20)=48.364$, p<0.0001 by item). Instruction type was a significant factor in a logistic regression of interpretation (coefficient: -1.206, p<0.03). Given the biases already noted for each coherence relation, the large number of Occasion relations in the 'What next' condition and the large number of Explanations in the 'Why' condition predict more Source interpretations in the 'Why' condition than in the 'What next' condition, as shown in Figure 4.
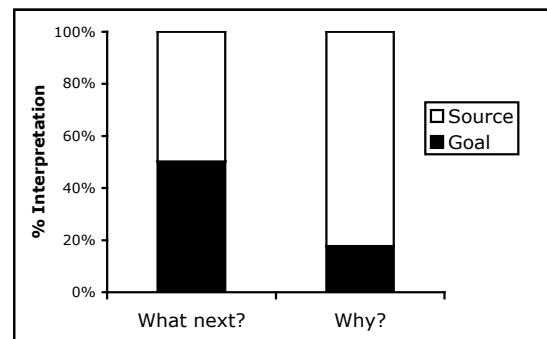

Figure 4: Effect of instruction on pronoun interpretation

---

[7] Again, the chi-squared test shows significant differences in the coherence distribution by instruction type (χ^2=734.122, p<0.0001).
[8] As before, all reported analyses represent a conservative treatment of the ambiguous data, excluding cases that were judged to be ambiguous by at least one judge (11% of total).

There were also significant interactions between instruction type and aspect (for pronoun interpretation, $F(1,40)=27.499$, $p<0.0001$ by subjects and $F(1,20)=41.187$, $p<0.0001$ by items, instruction×aspect coefficient: 2.022, $p<0.0001$; for modeling the distribution of coherence relations, instruction×aspect coefficient: 2.568, $p<0.0001$). These interactions can be seen in Figures 5 & 6.
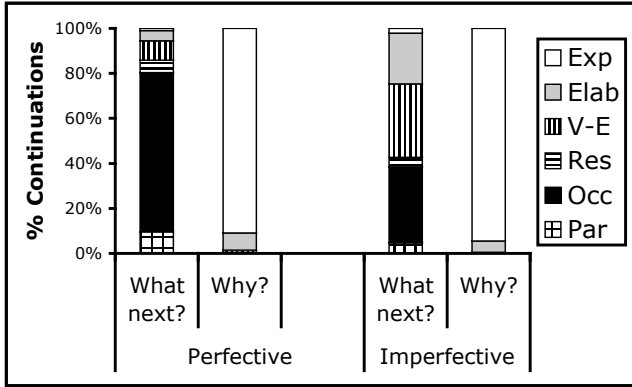

Figure 5: Effect of verbal aspect and instruction type on coherence relations
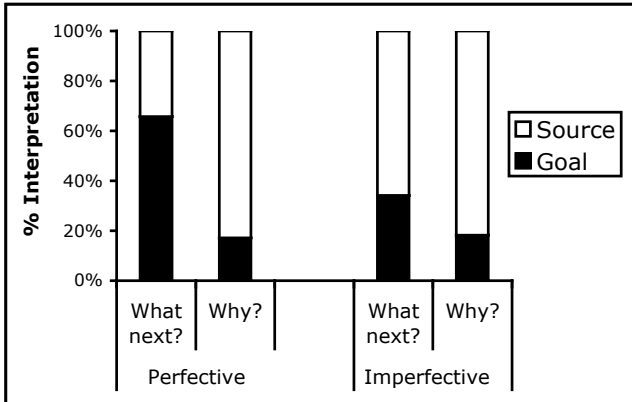

Figure 6: Effect of aspect and instruction on interpretation

Figure 5 shows how the effect of instruction type on the coherence distribution differs depending on the verbal aspect of the context sentence. The question 'Why' yielded predominantly Explanations across the board; the question 'What happened next', however, yielded mostly Occasions following perfective context sentences but a larger variety of relations following imperfective context sentences. This variation may stem in part from the lack of a salient endpoint in events described with imperfective aspect—making Occasion relations less appropriate and thereby yielding a wider range of alternative continuations.

These different coherence distributions impact the pronoun interpretation patterns, as shown in Figure 6. The question 'Why' yielded continuations that consistently referred to the Source regardless of aspect; the question 'What happened next' yielded pronoun interpretation patterns that differed by aspect. Responses to the question 'What happened next' following perfective context

sentences yielded a majority of Goal continuations, whereas the responses following imperfective context sentences yielded a majority of Source continuations. The point is that the interaction corresponds directly to the differences in the coherence distributions. In contexts in which the coherence distribution was more mixed ('What next' following imperfective context sentences), the interpretation pattern is also more mixed.

These subpatterns within the data show that different environments yield different distributions of coherence relations. No model of interpretation that ignores discourse coherence can account for these results, since the stimuli were identical between conditions. These results lay the groundwork for a predictive model in which knowing the coherence breakdown in a given context can help us predict the distribution of pronoun interpretations for these stimuli.

General Discussion

The results reported here support the development of a predictive model of pronoun interpretation that incorporates information about coherence relations. In what follows we illustrate how such a model could use interpretation biases calculated from one dataset to predict interpretation patterns observed in another dataset. The idea is that, given a coherence distribution for these stimuli, we can predict the distribution of pronoun interpretations because we already have estimates for the interpretation biases of each coherence relation.[9] Equation (8) calculates the percentage of Source resolutions in a context j using the interpretation biases calculated from a separate context i.

(8) Predict observed percentage of Source Resolutions in j with probabilities generated from i

$$\mathrm{Prob}_j(SR) = \sum_{coh} \%coh_j * p_i(SR \mid coh)$$
$$= \%elab_j * p_i(SR \mid elab) + \%exp_j * p_i(SR \mid exp) +$$
$$\%occ_j * p_i(SR \mid occ) + \%par_j * p_i(SR \mid par) +$$
$$\%res_j * p_i(SR \mid res) + \%v\text{-}e_j * p_i(SR \mid v\text{-}e)$$

In (8), we multiply the proportion of each coherence relation observed in a corpus j by the probability of observing a Source resolution given that coherence relation in corpus i, and then sum over all coherence relations. We can use equation (8) to estimate the percentage of Source resolutions in Experiment 2 ($\mathrm{Prob}_{EXP\text{-}2}(SR)$), using the coherence breakdown from Experiment 2 ($\%coh_{EXP\text{-}2}$) and the interpretation biases calculated from Experiment 1 ($p_{EXP\text{-}1}(SR|coh)$). Table 2 shows the result.

---

9 In neutral contexts, we would expect these stimuli to yield distributions of coherence relations that are correspondingly skewed toward the distributions previously reported, but as this paper shows, the distributions can be shifted by changing the context. One could imagine a context that yielded a corpus dominated by Elaboration relations, for example. In that case, the distribution of pronoun interpretations is predicted to be consistent with the interpretation bias observed for Elaboration relations in other contexts.

Table 2: Predicting % observed Source Resolutions (SR)

	Instruction	
	'What next'	**'Why'**
Observed SR across Exp 2	49.7%	82.3%
Predicted SR [probabilities generated from Exp 1]	45.4%	82.5%

As we have already mentioned, models that rely predominantly on surface-level cues (subject/first-mention preference: Gernsbacher & Hargreaves 1988, Crawley, Stevenson, & Kleinman 1990; grammatical parallelism: Sheldon 1974, see also Smyth 1994, Chambers & Smyth 1998; thematic role preference: Stevenson et al. 1994) do not similarly predict the variation between conditions. Our results suggest that even where pockets of data show evidence of grammatical-role, thematic-role, or event-level biases, these biases can and should be understood within a much richer model of discourse comprehension. These results also point to the importance of controlling for coherence factors when investigating surface-level or information-structural preferences (see Wolf, Gibson, & Desmet 2004; Kertz, Kehler, & Elman 2006).

Conclusions

Our story continuation experiments show that a shift in the distribution of coherence relations leads to a predictable, corresponding shift in pronoun interpretations even in cases in which stimuli are near-identical (Experiment 1) or entirely identical (Experiment 2). Our manipulations of object type and the experimental instructions are less important than the fact that one *can* alter the discourse environment and see a shift in pronoun interpretation. We sketched a predictive model of pronoun interpretation in which the salience of possible referents is linked to discourse-level expectations about the direction the discourse is likely to take.

Acknowledgements

References

Arnold, J. E. (2001). The effects of thematic roles on pronoun use and frequency of reference. *Discourse Processes, 31(2)*, 137-162.

Bates, D. (2006). Linear mixed model implementation in lme4. CRAN-R Package lme4.

Chambers, G. C. & Smyth, R. (1998) Structural parallelism and discourse coherence: A test of Centering Theory. *Journal of Memory and Language, 39*, 593-608.

Crawley, R., Stevenson, R., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research, 4*, 245-264.

Gernsbacher, M.A., & Hargreaves, D.J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, *27*, 699-717.

Hobbs, J. R. (1990) *Literature and Cognition.* CSLI Lecture Notes 21. Stanford, CA.

Kehler, A. (2002) *Coherence, reference, and the theory of grammar.* CSLI Publications, Stanford, CA.

Kertz, L., Kehler, A., & Elman, J. (2006) Grammatical and coherence-based factors in pronoun interpretation. In the Proceedings of the 28th Annual Conference of the Cognitive Science Society, Vancouver, July 26-29, 2006.

Rohde, H., Kehler, A., & Elman, J. (2006) Event Structure and Discourse Coherence Biases in Pronoun Interpretation. In the Proceedings of the 28th Annual Conference of the Cognitive Science Society, Vancouver, July 26-29, 2006.

Sheldon, A. (1974) The Role of Parallel Function in the Acquisition of Relative Clauses in English. *Journal of Verbal Learning and Verbal Behavior*, *13,* 272-281.

Smyth, R. H. (1994). Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research, 23*, 197-229

Stevenson, R., Crawley R., & Kleinman D. (1994). Thematic roles, focusing and the representation of events. *Language and Cognitive Processes*, *9*, 519–548.

Stevenson, R. 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) Coherence and pronoun resolution. *Language and Cognitive Processes*, *19(6)*, 665-675.

Appendix

Stimuli pairs: transfer verb & normal/abnormal object
1. bring a glass of water/a horse
2. carry a tray/a 40-pound turnip
3. chuck a wrench/an urn full of ashes
4. deliver a subpoena/four dozen lightbulbs
5. fax a resume/a picture of a panda
6. fling a frisbee/a stolen wallet
7. forward an email/a Viagra SPAM email
8. give a sweater/a hand grenade
9. hand a book/a medieval sword
10. kick a soccer ball/a pair of dirty sneakers
11. lob a football/a wet dog
12. mail a letter/a bloody meat cleaver
13. pass a sandwich/an empty jar of mayonnaise
14. roll a toy truck/a barrel of rancid pickles
15. send a fruitcake/a strand of dental floss
16. serve chili/stewed prunes
17. ship a package/a small lizard
18. throw a hat/a toilet plunger
19. toss a dish towel/a condom
20. transmit a message/a Shakespearean sonnet
21. wire money/one million dollars