

Dynamic Integration of Pragmatic Expectations and Real-World Event Knowledge in Syntactic Ambiguity Resolution

Klinton Bicknell (kbicknell@ling.ucsd.edu)

Department of Linguistics, UC San Diego
9500 Gilman Drive #108, La Jolla, CA 92093-0108 USA

Hannah Rohde (hannah@northwestern.edu)

Department of Linguistics, Northwestern University
2016 Sheridan Road, Evanston, IL 60208 USA

Abstract

This paper presents the results of a sentence completion study and a self-paced reading experiment involving a relative clause attachment ambiguity. Our results provide evidence that comprehenders resolve local structural ambiguity during incremental sentence processing by combining expectations about upcoming discourse continuations with their knowledge of typical relationships between event participants. We interpret these results as supporting an account in which comprehenders dynamically integrate cues from an in-principle unbounded range of sources during language comprehension.

Keywords: Syntactic ambiguity resolution; discourse processing; event knowledge.

Introduction

Psycholinguistic research shows that comprehenders rely on cues of many different types during language processing. We distinguish between two general explanations of the way that comprehenders store and use these cues. One possibility is that comprehenders appeal to a broad set of knowledge sources including discourse pragmatics and real-world event knowledge, integrating cues from this arbitrarily large set of sources as they incrementally process a sentence (an ‘arbitrary integration’ account). As Willits, Sussman, and Amato (2008) point out, however, another possibility is that comprehenders merely appear to be integrating cues from many sources, but instead are making use of simpler and easier to calculate statistical contingencies (correlated with knowledge sources such as event knowledge and discourse pragmatics) such as cooccurrence patterns (a ‘simple statistics’ account).

There is evidence from a variety of domains that comprehenders make use of both simple and more complex statistical models of linguistic events. For example, eye movements during reading can be significantly predicted using very simple probabilities such as the transitional probability from one word to the next (McDonald & Shillcock, 2003) as well as more complex probabilities such as the likelihood of a word given the preceding syntactic structure (Demberg & Keller, 2008). Also, the omission of optional material in language production (e.g., the rate of omission of optional *that*) has been found to vary with the predictability of the next word given the preceding syntactic context within the sentence (Levy & Jaeger, 2007).

There is also evidence that online processing is sensitive to a broader set of knowledge sources including discourse pragmatics and real-world event knowledge. For example, Simner

and Pickering (2005) show that expectations about upcoming discourse continuations are sensitive to the discourse’s causal structure. Comprehenders also appear to readily integrate their real-world knowledge of typical events during processing. Priming studies have shown facilitation between verbs and their typical agents, patients, instruments, and locations (e.g., *arresting~criminal* or *spoon~stirred*) within 250 ms (McRae, Hare, Elman, & Ferretti, 2005; Ferretti, Kutas, & McRae, 2007). Experiments using self-paced reading and event-related potentials have also shown that comprehenders’ use of event knowledge extends beyond single verb-to-participant dependencies and includes more complex knowledge about specific combinations of verbs, agents, patients, and instruments (Bicknell, Elman, Hare, & McRae, 2008; Matsuki, McRae, Tabaczynski, Elman, & Hare, 2008). Together, these studies suggest that comprehenders store a large body of knowledge about typical events and the relationships between event participants, and can deploy that knowledge rapidly during language comprehension.

In a recent paper, however, Willits et al. (2008) present evidence that some of the results mentioned above as support for the use of real-world event knowledge can also be modeled using simple word co-occurrence patterns. Specifically, they show that in Ferretti et al.’s (2007) priming study, the probability of the target occurring within 10 words of the prime in Wikipedia is higher in precisely those conditions that showed significant facilitation. Willits et al. argue that much of the evidence standardly taken to support the integration of a wide variety of knowledge sources can likewise be explained by simpler statistical models over linguistic events.

To test whether the simple statistics account is sufficient, we will present data that seem to demand an account in which comprehenders dynamically combine cues from real-world knowledge and discourse pragmatics during incremental sentence comprehension. In order to distinguish these two accounts, we refine the design of one of the existing studies that could perhaps be reinterpreted in terms of simple statistics (Rohde, Levy, & Kehler, 2008).

Rohde, Levy, & Kehler (2008)

Rohde et al. (2008) used contexts in which the syntactic attachment site of a relative clause is temporarily ambiguous.

- (1) John detests/babysits the children of the musician who...
- | | | | |
|---------------------------------------|--------------|-------------|----------------|
| <u>IC</u> | <u>NonIC</u> | <u>high</u> | <u>low</u> |
| a. ...are generally arrogant and rude | | | [high attach.] |
| b. ...is generally arrogant and rude | | | [low attach.] |

In these materials, a relative clause (RC) starting with *who* can be used to modify either of the noun phrases (NPs) in the preceding complex NP. The RC in (1a) is said to attach high to *the children*, the NP in the higher position in the syntactic structure; the RC in (1b) is said to attach low to *the musician*, the possessor NP within the the complex NP. Rohde et al. manipulated whether the matrix clause contained an *implicit causality* (IC) verb or not. IC verbs are a set of verbs that have been shown to pattern together in that they describe events in which one individual is implicated as central to the event's cause (e.g., *detest*, *adore*, *criticize*, *congratulate*; first discussed in Garvey & Caramazza, 1974). The IC verbs in their study were crucially all object-biased IC verbs – verbs like *detest* which describe an event whose cause is attributed primarily to the second NP, the direct object.

The results from Rohde et al.'s studies with sentence-completion and self-paced reading methodology show that low-attaching RCs are more frequent and easier to process following non-IC matrix clauses, but that the reverse holds in IC contexts: high-attaching RCs are more frequent and easier to process in contexts with IC matrix clause verbs.

Rohde et al. interpret this result as follows: for non-IC verbs, comprehenders make use of the default low-attachment preference that has been reported for English RCs (see review in Mitchell & Brysbaert, 1998, though cf. Gilboy, Sopena, Clifton, & Frazier, 1995 for discussion of variation in attachment preferences based on NP animacy and definiteness). For IC verbs, however, Rohde et al. argue that comprehenders use a combination of three discourse-level pragmatic factors: (i) an expectation given an IC verb that the next clause will provide an explanation (Kehler, Kertz, Rohde, & Elman, 2008), (ii) a bias in a subsequent explanation to re-mention the causally-implicated object NP of object-biased IC verbs (Garvey & Caramazza, 1974; McKoon, Greene, & Ratcliff, 1993; *inter alia*), and (iii) the knowledge that an RC can be used to provide an explanation. The use of the term 'pragmatic factors' here is intended to include those cues that guide the establishment of discourse coherence and coreference.

Applied to example (1), Rohde et al.'s account maintains that encountering the verb *detest* increases comprehenders' expectations for an upcoming explanation (one that will explain *why* John detests the children of the musician), whereas encountering the verb *babysit* creates no such expectation. At that point, if comprehenders are implicitly aware that explanations following a verb like *detest* are likely to be about the object NP (*the children*) and they are implicitly aware that the upcoming RC can be used to provide such an explanation, then they will prefer RCs that attach to *the children* more so than they will following a non-IC verb like *babysit*. Crucially, the object NP *the children* occupies the high-attachment position, so that the bias to provide an explanation centered

around *the children* following *detest* but not *babysit* yields attachment preferences that differ by verb type.

From this result, it appears that comprehenders are combining several discourse-level biases and expectations as they make their attachment decision mid-sentence. This study, however, could also be explained by a simpler statistical model. Just as comprehenders appear to track the subcategorization probabilities of different specific verbs (Trueswell, Tanenhaus, & Kello, 1993), they may keep track of the probabilities of high- versus low-attaching RCs following individual verbs, the verb-specific attachment height model. It is certainly possible that, given the inferencing described above, there is a statistical pattern present in language data in which more high attaching RCs follow IC verbs than non-IC verbs, so it is plausible that language users could make use of that. The next section describes how we refine the design of Rohde et al. (2008) to tease apart these two accounts.

Comparing the Two Accounts

The intuition behind our refinement is that since the verb-specific attachment height model predicts object-biased IC verbs to have a consistent bias for high-attaching RCs, then these attachment preferences should not interact with other cues drawn from knowledge beyond the immediate sentence, including, for example, cues related to the specific individuals mentioned in the complex NP and the real-world relationship that holds between those individuals. On the other hand, if, as in the arbitrary integration account, comprehenders are integrating several pragmatic and real-world knowledge cues together for disambiguation, then the bias for high-attaching RCs following IC verbs should interact with other information sources.

In our materials, we use the basic design from Rohde et al. (2008) with IC and non-IC verbs followed by temporarily ambiguous RCs, but we add one more factor: the 'responsibility relationship' that stereotypically holds between the individuals referenced by the two NPs:

- (2) John criticizes_{IC}/resembles_{non-IC} ...
- | | |
|--|-------------------|
| a. the orthodontist of the kids who... | [non-responsible] |
| b. the father of the kids who... | [responsible] |

At an intuitive level, a father is responsible for his kids' behavior in a way that an orthodontist is not. In Rohde et al.'s materials, the responsibility relation between the individuals in the complex NP was not tightly controlled, but most items exhibited a non-responsible relation like the one in (2a). If, as under the arbitrary integration account, comprehenders are combining pragmatic and real-world knowledge during incremental processing, we predict this sort of responsibility relation to modulate the effect of IC verbs.

The crucial intuition is that the reason that someone criticizes an orthodontist of some kids is likely to center around the orthodontist's behavior, whereas the reason that someone criticizes the father of some kids may instead center around the kids' behavior precisely because the father is, to some

extent, responsible for kids' behavior. More specifically, according to Rohde et al.'s interpretation of their results (as in the description given above), encountering the IC verb *criticize* increases expectations for an upcoming explanation (one that will explain *why* John criticizes the orthodontist/father), whereas encountering the verb *resemble* creates no such expectation. Furthermore, just as under Rohde et al. (2008)'s account, comprehenders must be implicitly aware that an RC can be used to provide an explanation.

The responsibility relationship gives rise to the difference in predictions in the final part of the reasoning. When no responsibility relationship exists (2a), the predictions replicate Rohde et al.: An RC providing an explanation about why the orthodontist is criticized is likely to attach high to *orthodontist*. On the other hand, when a responsibility relationship does exist (2b), an RC providing an explanation for why the father is criticized could still attach high to *father*, but additionally, it could attach low to *kids*, since a father is somewhat responsible for his kids' behavior (e.g., *John criticizes the father of the kids who are often seen playing in traffic*). Thus, our predictions are that, for non-responsible cases, IC verbs should increase the proportion of high attachments relative to non-IC verbs, but for responsible cases, IC verbs will not increase the proportion of high attachments by as much – and in fact may even decrease them. Verification of these predictions would represent evidence that comprehenders can combine knowledge from a range of sources together during incremental comprehension, and are unlikely to be relying solely on linguistic cooccurrence generalizations.

The remainder of this paper describes two experiments that test these predictions. We present the results of a norming study to evaluate the responsibility relations between pairs of individuals. We then present the results of an offline sentence-completion experiment that demonstrates that RC attachment biases differ, as per the above predictions, in contexts with responsible and non-responsible pairs. Finally, we show that the attachment biases revealed in the offline study influence comprehenders' online processing in a self-paced reading experiment.

Norming Study: Responsibility Relations

The norming study was conducted to probe comprehenders' intuitions about the strength of the responsibility relation between individuals in a set of NP pairs.

Methods

Participants 22 students from UC San Diego participated for course credit. All were monolingual English speakers.

Materials We selected 44 NPs (e.g., *the kids*) and 2 related NPs for each, one predicted to be responsible for first NP (*the father*) and one not (*the orthodontist*). These were combined with object-biased IC verbs into plausibility questions, shown for one item in (3a-b).

(3) a. [*Non-Responsible*] Imagine an orthodontist of

some kids. How plausible would it be for someone to criticize the orthodontist for something the kids did?

b. [*Responsible*] Imagine a father of some kids. How plausible would it be for someone to criticize the father for something the kids did?

We selected two IC verbs for each item, and half the participants saw the item with a negative polarity IC verb (*criticize*) and half with a positive polarity IC verb (*congratulate*).

Procedure We administered a web-based questionnaire that contained questions like those in (3a-b). Participants rated the plausibility of the scenarios on a scale from 1 to 7 with 1 being least plausible and 7 being most plausible. Participants responded to all 44 items in both the responsible condition and the non-responsible condition, all in a separately randomized order, for a total of 88 responses.

Results and Discussion

For each item, we calculated the mean rating for each combination of IC verb (positive/negative polarity) and responsibility relation (responsible/non-responsible), yielding 4 scores per item. We then computed the mean effect size of responsibility for each item with each verb by subtracting the rating for the non-responsible condition from that for the responsible. We used these scores to select the 40 strongest items (and best polarity for each item) for use in the sentence-completion experiment. For the 40 best items, the mean plausibility rating for the non-responsible condition was 2.1 (*SE* 0.1), and for the responsible condition it was 5.3 (*SE* 0.1), demonstrating that comprehenders are aware of and sensitive to the responsibility differences in these materials.

Sentence Completion Experiment

We conducted a sentence-completion experiment varying verb type (IC/non-IC) and NP-pair responsibility (responsible/non-responsible) in order to test the effects of discourse-level biases and real-world event knowledge on RC attachment preferences. The verb-specific attachment height model predicts no interaction of NP-pair responsibility and verb type in attachment preferences, and the only anticipated effect is an effect of verb type, replicating Rohde et al.. (Neither account makes any predictions as to whether or not there will be a main effect of responsibility.) The arbitrary integration account, however, predicts an interaction whereby non-responsible NP pairs yield more high-attaching RCs following IC verbs than non-IC verbs (replicating Rohde et al.), but this difference is reduced for responsible NP pairs.

Methods

Participants 44 students from UC San Diego participated for course credit. All were monolingual English speakers.

Materials We combined the noun-noun-verb triplets obtained from the norming study with a proper name and a non-IC verb to create sentence-completion prompts like (4).

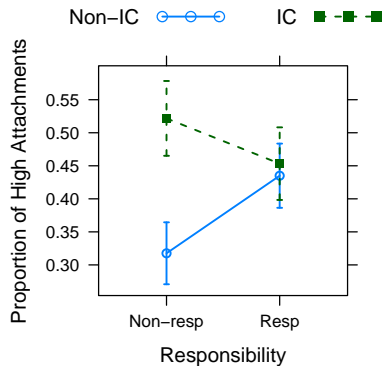


Figure 1: Proportion of high attachments by verb type and responsibility. Error bars show standard error of the mean.

- (4) John criticizes_{IC}/resembles_{NonIC} the orthodontist_{NonResp}/father_{Resp} of the kids who _____

These 40 items were combined with 60 fillers, which contained a mix of embedded clauses, subordinate clauses, RC prompts, unambiguously attached RCs following complex NPs, and other *wh*- words as prompts.

Procedure The sentence-completion experiment was conducted using a web-based interface. Each participant saw each item only once. Participants were asked to imagine a natural completion to the sentence and write the first completion that came to mind.

Results and Discussion

Two judges, the authors of this paper, assessed the elicited RCs for their attachment height. Attachment heights were annotated as either high, low, or ambiguous, and were only assigned the high/low labels when the intent of the writer was clear, either through number agreement on the verb or through event plausibility.

After excluding 18 ungrammatical/nonsensical completions as well as the 35.5% of completions which were not disambiguated through syntactic number or event plausibility, we analyzed the remaining completions for effects of verb type and the responsibility relation on RC attachment. The data are shown in Figure 1. As predicted, the non-responsible condition replicated Rohde et al.’s results, showing the default low-attachment preference for non-IC verbs (31.8% high attachments) and an increase in high attachments for IC verbs (52.2% high attachments). In the responsible condition, however, the proportion of high attachments was similar for IC verbs and non-IC verbs (IC: 45.3%; non-IC: 43.5%).

We performed two statistical analyses on this data, conducting a 2×2 ANOVA (responsibility \times verb type) on the arcsine-transformed percentages as well as fitting a generalized linear mixed-effects regression with a logit link function (Agresti, 2002; Bates & Sarkar, 2008) to the raw attachment data, allowing for the use of random intercepts by participant and by item. Results for the latter model will be

given as a Wald Z-score and associated p -value. The results showed a significant main effect of verb type [$F_1(1, 43) = 14.4, p < .001$; $F_2(1, 39) = 6.1, p < .05$; $Z = 5.2, p < .001$], driven by the non-responsible condition which replicates Rohde et al.. Crucially, there was also a significant interaction between responsibility and verb type [$F_1(1, 43) = 11.5, p < .01$; $F_2(1, 39) = 5.6, p < .05$; $Z = -4.1, p < .001$], showing that as predicted the difference in attachment preferences for IC and non-IC verbs is reduced in contexts with responsible NP pairs. This result thus supports the arbitrary integration account, demonstrating that comprehenders combine pragmatic information with real-world knowledge to make syntactic attachment decisions.

Self-Paced Reading Experiment

In order to test whether the biases observed in the offline sentence-completion experiment influence comprehenders’ online sentence processing, we conducted a $2 \times 2 \times 2$ self-paced reading experiment varying verb type, responsibility, and attachment height. We adapted the sentence completion materials to create sentences with a disambiguating verb that signaled the RC attachment height. We measured reading times at the critical disambiguating region and one region after, testing for a three-way interaction between attachment height, verb type, and responsibility. A two-way interaction alone between attachment height and verb type would support the verb-specific attachment height model whereby high-attaching RCs are easier to process in contexts with object-biased IC verbs than non-IC verbs, regardless of the responsibility relation. A three-way interaction, on the other hand, would support the arbitrary integration account whereby responsible NP pairs reduce the facilitation provided for high-attaching RCs by IC verbs.

Methods

Participants 40 students from UC San Diego participated for course credit. All were monolingual English speakers.

Materials We used all 40 prompts from the sentence-completion experiment and added RC endings that signaled attachment height through number agreement on the embedded RC verb:

- (5) John criticizes_{IC}/resembles_{NonIC} the orthodontist_{NonResp}/father_{Resp} of the kids who is_{high}/are_{low} often heard yelling loudly.

High-attaching RCs were signaled with a singular verb (*is/was/has*) and low-attaching RCs were signaled with a plural verb (*are/were/have*). The next word after the RC verb was a neutral adverb and was the same in all conditions. The 40 target sentences were mixed with 60 filler sentences, created by adding completions to the 60 sentence-completion fillers.

Procedure Participants read the sentences on a computer screen using a word-by-word, non-cumulative, self-paced moving-window reading paradigm (Just, Carpenter, & Wool-

ley, 1982). Each trial began by displaying the entire target sentence on screen with dashes replacing all the letters. Participants pressed a button on a game controller to reveal the next word, which simultaneously caused the previous word to change back to dashes. During this procedure, the time between the presentation of a word and the button press was recorded. After each sentence, a comprehension question was displayed. Participants used two other buttons on the controller to indicate their answer. They were instructed to read as naturally as possible, and to slow down if they found themselves unsure of an answer to a comprehension question.

Results and Discussion

Any reading time for which a participant spent longer than 3.5 seconds reading a single word was excluded from the analysis (102 reading times total across fillers and experimental items). Next, in order to minimize the effects of word length differences across conditions and of participant reading rates, a regression was performed on the data for each participant, predicting reading time from word length in characters. This process used all words from both filler and experimental trials. The values predicted by the regressions were subtracted from the actual reading times to produce residual reading times. Residual reading times that were more than 2.5 standard deviations from the mean for each combination of position and condition were excluded from further analysis. This process affected less than 1% of the data.

Comprehension question accuracy Overall, comprehension question accuracy on fillers and target items was 89% (SE 0.7%), and no single participant was below 78%. We performed two analyses on the experimental items: A three-way ANOVA (responsibility \times verb type \times attachment) on the arcsine-transformed accuracy data, and a generalized linear mixed-effect regression on the raw accuracy data. Neither analysis showed any main effect or interaction.

Reading times We performed two analyses on the residual reading times for the critical verb and the immediately following adverb: a three-way ANOVA and a linear mixed-effect regression (Pinheiro & Bates, 2000), which allowed for the inclusion of random intercepts for participants and items. For the latter model, we report a t value and a p -value produced by Markov chain Monte Carlo sampling (Baayen, In Press).

The residual reading times for the critical verb and following adverb are given in Table 1.¹ On the verb, there was a main effect of responsibility [$F_1(1, 39) = 5.1, p < .05; F_2(1, 39) = 5.8, p < .05; t = -1.9, p < .05$] such that the responsible condition was faster than the non-responsible. This suggests that comprehenders had an easier time with the responsible nouns, perhaps indicating that, e.g., there are closer semantic connections between *father* and *kids* than *orthodontist* and *kids*. The ANOVA also showed a significant effect of attachment [$F_1(1, 39) = 5.4, p < .05; F_2(1, 39) = 7.2, p < .05$], indicating that low attachment was preferred

Table 1: Residual reading times (and SE) for the Verb and Adverb regions by verb type, attachment, and responsibility.

	Non-IC		IC	
	High	Low	High	Low
Verb/Non-resp.	28 (13)	-13 (13)	24 (14)	-11 (11)
Verb/Resp.	-2 (11)	-1 (14)	-13 (13)	-19 (9)
Adverb/Non-resp.	-29 (19)	-73 (15)	-58 (20)	-54 (15)
Adverb/Resp.	-75 (12)	-72 (15)	-44 (15)	-72 (16)

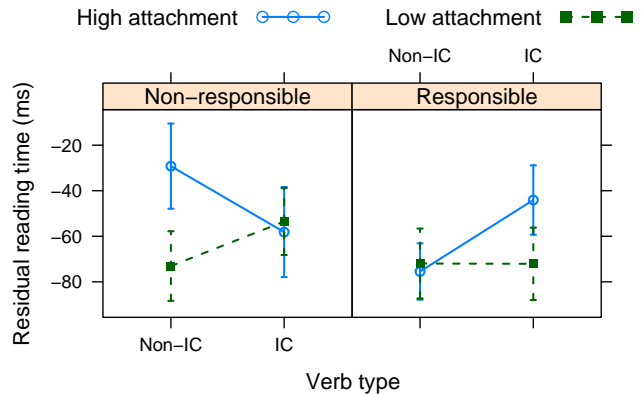


Figure 2: Residual reading times for the Adverb region by responsibility, verb type, and attachment. Error bars show standard error of the mean.

overall. Finally, there was a significant interaction between responsibility and attachment [$F_1(1, 39) = 4.1, p < .05; F_2(1, 39) = 6.0, p < .05; t = 2.4, p < .05$] such that while low attachments were about 40 ms faster than high attachments in the non-responsible condition, the two attachment heights were quantitatively similar for responsible contexts. While we did not specifically predict this effect, one explanation may be that the probability of seeing a verb in this position varies across the conditions. Such a suggestion would, however, require further testing.

Residual reading times for the adverb region are plotted in Figure 2. On the adverb, there were no significant main effects or two-way interactions. There was, however, the predicted three-way interaction [$F_1(1, 39) = 5.1, p < .05; F_2(1, 39) = 4.4, p < .05; t = 2.0, p < .05$]. Looking at Figure 2, this three-way interaction can be interpreted as follows: for non-responsible noun pairs, IC verbs make high attachments easier, as found by Rohde et al. (2008); for responsible noun pairs, however, IC verbs actually serve to make high attachments harder. That is, whereas hearing about criticism of an orthodontist creates expectations to hear more information about the orthodontist (relative to *kids*), hearing about criticism of a father actually creates expectations to hear more information about his kids. This online result thus mirrors the offline sentence completions in supporting the arbitrary integration hypothesis; that is, it shows that comprehenders dynamically combine pragmatic and real-world knowledge to

¹All patterns reported look qualitatively similar with raw RTs.

resolve syntactic ambiguity incrementally.

Conclusion

This paper presents a set of experiments testing the predictions of two general explanations for results in language processing research: simple statistics and arbitrary integration. Together the experiments provide evidence for a case in which comprehenders appear to track arbitrary real-world event information such as the responsibility relationship between two individuals (norming study), that they integrate this information along with verb-driven pragmatic cues to help resolve syntactic ambiguity (sentence-completion experiment), and lastly that they perform this integration of real-world and pragmatic knowledge in their incremental sentence processing (self-paced reading experiment). The results support the view that syntactic processing reflects a dynamic combination of discourse-driven biases with real-world knowledge as idiosyncratic as the responsibility relationship between two individuals.

Of course, one could build a more complex statistical model of this phenomenon than the verb-specific attachment height model in order to predict the interactions we found. The model would need to track the probability of high- vs. low-attachment after each combination of a particular matrix verb with two particular NPs. At this point, however, it is not clear that this model would represent a simpler account of the data. Thus, we maintain that our results are best interpreted as supporting the arbitrary integration account.

This claim fits within a growing body of literature suggesting that language comprehension requires the integration of multiple cues from the linguistic context and from comprehenders' real-world event knowledge. To take one example, work by Desmet, de Baecke, and Brysbaert (2002) establishes that RC attachment ambiguity is sensitive to discourse-level linguistic properties of the relevant NPs (in their case, the NPs' definiteness in contexts in which the referent is or is not uniquely identifiable); our work can be seen as extending this result by showing that RC attachment ambiguity is sensitive to a combination of linguistic cues (in our case, cues that are driven by the verb) and non-linguistic real-world knowledge cues (here, the responsibility relationship between individuals). Experimental work is consistently finding that incremental sentence processing is sensitive to more and more complex interactions between cues. This implies that comprehenders are quite skilled at combining multiple cues and that our models must accommodate these complex interactions between many different domains of knowledge.

Acknowledgments

This paper benefitted greatly from feedback and discussion from Roger Levy. The research was supported by NIH Training Grant T32-DC000041 from the Center for Research in Language at UC San Diego to the first author and by an Andrew W. Mellon Postdoctoral Fellowship to the second author.

References

- Agresti, A. (2002). *Categorical data analysis* (Second ed.). New York, NY: John Wiley & Sons.
- Baayen, R. (In Press). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.
- Bates, D. M., & Sarkar, D. (2008). *lme4: Linear mixed-effects models using Eigen and Eigen++*. R package version 0.999375-23.
- Bicknell, K., Elman, J. L., Hare, M., & McRae, K. (2008). Online expectations for verbal arguments conditional on event knowledge. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2220–2225). Washington, D.C.: Cognitive Science Society.
- Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2), 193–210.
- Desmet, T., de Baecke, C., & Brysbaert, M. (2002). The influence of referential discourse context on modifier attachment in Dutch. *Memory and Cognition*, 30, 150–157.
- Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(1), 182–196.
- Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, 5(3), 459–464.
- Gilboy, E., Sopena, J., Clifton, C., & Frazier, L. (1995). Argument structure and association preferences in Spanish and English complex NPs. *Cognition*, 54, 131–167.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111(2), 228–238.
- Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, 25, 1–44.
- Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in Neural Information Processing Systems 19* (pp. 849–856). Cambridge, MA: MIT Press.
- Matsuki, K., McRae, K., Tabaczynski, T., Elman, J. L., & Hare, M. (2008). *Conceptual and structural event-based expectancy generation in language comprehension*. Poster presentation at the 2008 CUNY Conference on Human Sentence Processing, Chapel Hill, NC, 13–15 March 2008.
- McDonald, S. A., & Shillcock, R. C. (2003). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, 43, 1735–1751.
- McKoon, G., Greene, S., & Ratcliff, R. (1993). Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology*, 19, 1040–1052.
- McRae, K., Hare, M., Elman, J. L., & Ferretti, T. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, 33(7), 1174–1184.
- Mitchell, D., & Brysbaert, M. (1998). Challenges to recent theories of crosslinguistic variation in parsing: evidence from Dutch. *Syntax and Semantics*, 31, 313–335.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed effects models in S and S-Plus*. New York: Springer Verlag.
- Rohde, H., Levy, R., & Kehler, A. (2008). *Implicit causality biases influence relative clause attachment*. Presented at the 21st Annual CUNY Conference on Human Sentence Processing, Chapel Hill, NC, March 13–15, 2008.
- Simner, J., & Pickering, M. (2005). Planning causes and consequences in discourse. *Journal of Memory and Language*, 25, 226–239.
- Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 528–553.
- Willits, J. A., Sussman, R. S., & Amato, M. S. (2008). Event knowledge vs. verb knowledge. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2227–2232). Austin, TX: Cognitive Science Society.