The Pragmatics of Characters' Mental Perspectives in Pronominal Reference Resolution

Tiana V. Simovic
University of Toronto
tiana.simovic@mail.utoronto.ca

Craig G. Chambers
University of Toronto
craig.chambers@utoronto.ca

Abstract

To date, cognitive models of pronoun resolution have primarily focused on how fairly shallow discourse-level and lexical cues yield the appropriate interpretation, despite classic work in computational linguistics emphasizing the importance of situation-specific pragmatic reasoning. We explore the latter in two studies of human judgments, which highlight the striking robustness of these pragmatic processes.

1 Introduction

Models of pronoun resolution are typically built around comparatively "shallow" heuristics such as discourse-level cues (e.g., first-mention cues, focus tracking, "Centering" (Grosz et al., 1995)) and lexical cues derived from semantic aspects of the verb (e.g., so-called "implicit causality"). These cues are readily implemented in both small- and large-scale models and have been pursued with the hope that these models would achieve high accuracy without the need to incorporate rich knowledge postulates and pragmatic reasoning. Work in psycholinguistics has reflected this same focus, with the majority of studies exploring how discourse-level and lexical cues guide human intuitions about referent identity (Kaiser and Fedele, 2019). This work has often concluded that discourse/lexical cues provide a kind of rapid default interpretation, as reflected by statistical tendencies in human judgments. Interestingly, this shared approach fails to capture many important insights from classic work in computational linguistics, which highlighted how situation-specific pragmatic reasoning is essential for resolving pronouns in many circumstances (e.g., Winograd, 1972; Hobbs, 1979; Hobbs et al., 1993). Given that cases involving situational reasoning are often described as challenging for computational models (e.g., Levesque et al., 2012; Sakaguchi et al., 2021), and are often incompatible with the solution yielded by default bias, it is possible that they are also difficult for humans to interpret. This would be reflected in less robust judgments compared to cases where pragmatic inferencing is not necessary for accurate identification of the intended referent. However, consider the following example from Jones and Bergen (2021):

(1) a. When the vase fell on the rock, it broke.
b. When the rock fell on the vase, it broke.

A resolution account based on shallow cues would predict that the pronoun *it* should resolve to the subject antecedent in both (1a-b). However, Jones and Bergen found that human readers judge the object-position antecedent (*vase*) in (1b) as the intended referent 95% of the time (e.g., despite that antecedent's status as the second-mentioned and therefore less "focal/centered" entity). This finding highlights how readers draw on world knowledge – something that continues to be difficult to integrate into current models of anaphora resolution (Richard-Bollans et al., 2018).

The present study extends the psycholinguistic work on inference in pronoun resolution by exploring how another form of world knowledge, namely mentalizing and perspective-taking about story characters, guides human pronoun resolution. This line of work provides challenging test cases for state-of-the-art computational models of coreference resolution in English.

2 Experiment 1: Subject Pronoun judgment Task

The first experiment (54 adult participants, M_{age}=34.54 years, SD_{age}=12.8, recruited from Prolific [www.prolific.com]; 24 critical trials) focused on *subject*-position pronouns using short sentences like in (2):

(2) a. Madeline told Anna that she remembers when the lecture starts.

Proceedings of the Sixth Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC 2023), pages 48–50
December 6–7, 2023

b. Madeline asked Anna if she remembers when the lecture starts.

We predicted that a character *telling* an interlocutor about the information expressed in the subordinate clause should lead readers to interpret the pronoun as coreferring with the main-clause subject, whereas *asking* should entail main-clause object selections. This is because (in relation to the examples in (2)) we do not normally tell people what they remember (conversational contributions should be informative, cf. Grice, 1975), and we do not normally ask people what we ourselves remember (e.g., Brown-Schmidt et al., 2008).

The results overwhelmingly supported these predictions: Participants chose the "perspectivally congruent" antecedent 99.8% of the time. The robustness of this judgment is striking relative to the strength of the patterns observed in computational and psycholinguistic studies exploring the effectiveness of superficial discourse/lexical cues (e.g., Tetreault, 2001; Kehler and Rohde, 2013). Further, there was no order-of-mention bias (which would predict more pronounced effects for *tell*, where the antecedent is the first-mentioned character). Specifically, readers picked the subject antecedent 99.7% of the time in the *tell* sentences and the object antecedent 99.8% of the time in the *ask* sentences. This illustrates that the pragmatic reasoning in question completely overrules the influence of canonical discourse effects related to order-of-mention, which is the pattern otherwise predicted in Centering and most other focus-based models.

3 Experiment 2: Object Pronoun judgment Task

To ensure the patterns are not due to readers drawing on statistical patterns regarding how arguments in a clause containing *tell* or *ask* are linked to a subsequent subject pronoun, we conducted the same experiment with *object* pronouns.

The experiment (54 adult participants, M_{age}=33.83 years, SD_{age}=13.43, recruited from Prolific [www.prolific.com]; 24 critical trials) was the same as Experiment 1, except that we now used sentences with object-position pronouns as in (3):

(3) a. Nina told Mary that modern art interests her more than classics.
b. Nina asked Mary if modern art interests her more than classics.

The results reflected the same reasoning-driven patterns as in Experiment 1, with the perspectivally-congruent antecedent selected 99.4% of the time (99% subject antecedent selection in *tell* sentences, and 99.7% object antecedent selection in *ask* sentences). The *ask* case result again demonstrates the apparent dominance of pragmatic reasoning over discourse- and structural-based cues in pronoun resolution.

4 Discussion

The judgment tasks showed extremely robust effects of perspectival inference on pronoun interpretation, suggesting that discourse biases are completely overruled by pragmatic reasoning, consistent with the findings from Jones and Bergen (2021). However, an alternative explanation that might be compatible with minimal use of world knowledge and pragmatic reasoning is that readers are drawing on stored "constructions" of some kind (Goldberg, 1995), as illustrated in (4) and (5):

(4) NP_1 told NP_2 [that] ... $PRONOUN_1$ ...

(5) NP_1 asked NP_2 [if] ... $PRONOUN_2$ ...

However, when we begin extending our consideration of these "perspective" discourses further, it becomes apparent that changes to other aspects of the sentences can strongly shift intuitions:

(6) a. Jane, who noticed it was 12:30 PM, was walking with her good friend Hana.
b. Jane, who is unfamiliar with Japanese currency, was talking to her tour guide, Hana.
c. Jane asked Hana if she had enough cash to buy a sandwich.

(7) a. Susan asked Molly if she likes pie.
b. Little Sue asked her mom if she likes pie.

(8) Max told Gerald that he had lint on the back of his coat.

Our preliminary work shows that when readers are shown either (6a) or (6b) and then prompted for judgments about whom the pronoun *she* refers to in (6c), readers shift from choosing the subject antecedent 12.5% of the time in (6a) to 100% of the time in (6b), suggesting the context provided in (6b) overrides typical *ask* selections by encouraging a different understanding of which character possesses the relevant knowledge (epistemic authority) for the question under discussion. Similarly, although (7a-b) share the same structure and predi-

cates, the understood antecedents clearly shift. The pronoun in sentence (7a) should again follow the pattern we found with our *ask* materials, however, (7b) suggests that Little Sue, who is presumably a child, can be asking her mom whether she herself likes to eat pie. Finally, (8) seems to demonstrate the opposite pattern of our *tell* materials, where the pronoun intuitively corefers with *Gerald*, the object antecedent of the sentence, as this character would be most likely to possess the knowledge in question. Given these examples, which all reflect intelligent perspective reasoning, it is unlikely that reliance on some abstract form of verb-specific frames underlies the observed patterns in the experiments reported above.

In summary, the findings highlight the cost of not including world knowledge and reasoning (cf. Grice, 1975) into current models of pronoun resolution and also underscore the benefit of expanding the standard stock of test cases when creating performance benchmarks for automated systems (Byron, 2003; Webster et al., 2018). We are currently assessing human judgments for cases like (6) and (7) to further test the importance of perspective cues and substantiate the account advanced above that readers' selection patterns are a result of intelligent reasoning and world knowledge rather than a reliance on shallow cues like sentence frames. We hope our work will inform the design of future benchmarks and computational models of anaphora resolution.

Limitations

A limitation of our work is that we only tested a narrow range of experimenter-constructed materials. Future work should extend this analysis to a wider range of materials, including similar cases found in naturalistic corpora.

Further, this work could be extended to languages beyond English that have different anaphoric patterns, such as the occurrence of zero pronouns in Japanese.

References

Sarah Brown-Schmidt, Christine Gunlogson, and Michael K. Tanenhaus. 2008. Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, 107(3):1122–1134.

Donna Byron. 2003. Annotation of pronouns and their antecedents: A comparison of two domains. *Technical Report 703, University of Rochester*.

Adele E. Goldberg. 1995. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

Herbert P. Grice. 1975. Logic and conversation. In *Speech Acts*, pages 41–58. Brill.

Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

Jerry R. Hobbs. 1979. Coherence and coreference. *Cognitive Science*, 3(1):67–90.

Jerry R. Hobbs, Mark E. Stickel, Douglas E. Appelt, and Paul Martin. 1993. Interpretation as abduction. *Artificial Intelligence*, 63(1–2):69–142.

Cameron R. Jones and Benjamin Bergen. 2021. The role of physical inference in pronoun resolution. In *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*, volume 43.

Elsi Kaiser and Emily Fedele. 2019. Reference resolution: A psycholinguistic perspective. In Jeanette Gundel and Barbara Abbott, editors, *The Oxford Handbook of Reference*. Oxford University Press.

Andrew Kehler and Hannah Rohde. 2013. A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, 39(1-2):1–37.

Hector Levesque, Ernest Davis, and Leora Morgenstern. 2012. The Winograd Schema Challenge. In *13th International Conference on the Principles of Knowledge Representation and Reasoning*, pages 552–561.

Adam L. Richard-Bollans, Lucía Gómez Álvarez, and Anthony G. Cohn. 2018. The role of pragmatics in solving the Winograd Schema Challenge. In *Proceedings of the 13th International Symposium on Commonsense Reasoning (Commonsense 2017)*. CEUR Workshop Proceedings.

Keisuke Sakaguchi, Ronan Le Bras, Chandra Bhagavatula, and Yejin Choi. 2021. Winogrande: An adversarial Winograd Schema Challenge at scale. *Communications of the Association for Computing Machinery*, 64(9):99–106.

Joel R. Tetreault. 2001. A corpus-based evaluation of Centering and pronoun resolution. *Computational Linguistics*, 27(4):507–520.

Kellie Webster, Marta Recasens, Vera Axelrod, and Jason Baldridge. 2018. Mind the gap: A balanced corpus of gendered ambiguous pronouns. *Transactions of the Association for Computational Linguistics*, 6:605–617.

Terry Winograd. 1972. Understanding natural language. *Cognitive Psychology*, 3(1):1–191.