

Ideology as Predictive Bias: Political Beliefs Modulate Discourse Relation Interpretation

Abhishek Verma, Frances Yung, Vera Demberg

Discourse relations (such as causality, contrast, and concession) are often unmarked by explicit connectives, requiring readers to infer the intended relation from context. Prior implicit discourse annotation work shows that disagreements arise when annotating implicit discourse relations, not only among crowdsourced non-expert annotators but among trained expert annotators as well [1][2][4]. This inferential process is not purely linguistic: research in discourse psychology suggests that comprehension is shaped by subjective knowledge and ideological schemata [3]. In this work, we explore political ideology as a systematic yet understudied source of predictive bias in discourse processing.

We investigate whether political beliefs modulate discourse expectations during implicit discourse comprehension, operationalized through a crowdsourced connective insertion task. Participants are presented with implicit discourse pairs (sentence pairs lacking an explicit connective) constructed around politically salient topics, and asked to insert an explicit connective by dragging it between sentences from a provided set. Consider the pair:

The workplace diversity training included mandatory pronoun-sharing exercises. Several employees declined to participate.

Such pairs are genuinely ambiguous, supporting multiple interpretations (causal, concessive, contrastive, etc.). A participant might insert a causal connective such as “as a result” or “therefore”, following the reasoning that since the training was mandatory and perceived as an overreach, declining to participate was a natural, expected outcome of such an imposition. Alternatively, a participant might insert a concessive connective such as “however” or “but”, treating non-participation as surprising: despite the exercises being standard and inclusionary in intent, some employees still refused, contrary to what one might expect. Which of these readings feels more natural is not arbitrary — we argue it reflects the reader's prior political beliefs and values. A reader who views mandatory diversity initiatives skeptically may find the causal framing intuitive (e.g., a conservative in the current U.S. political climate), while a reader who sees such initiatives as reasonable norms may find non-participation unexpected, and thus concessive (e.g., a liberal in the current U.S. political climate). We therefore propose that differences in political ideology are a key factor shaping which connective and discourse relation readers infer, and hypothesize that individuals with different political orientations will systematically differ in their inferred relations and connective insertion preferences (H1).

We create the experimental items (two-sentence discourse pairs), built around politically salient topics in the current U.S. political climate, designed to elicit ideologically divergent interpretations. We follow a general pattern where one sentence describes an ideologically salient policy, norm, or institutional action, and the other describes an interpretation or behavioral reaction to that action by a neutral institution (or person, e.g. “several employees”); candidate pairs are generated via LLM prompting, then manually revised and filtered for quality. To further select stimuli from this set, we conducted a coherence rating pre-test in which participants rated explicit causal and concessive versions of each candidate discourse pair, with each participant seeing only one version per item. Items were retained if both versions met a minimum coherence threshold, confirming that each version was independently plausible, and if the two

versions showed comparable overall plausibility relative to one another, confirming genuine interpretive ambiguity between the causal and concessive readings. This procedure yielded 22 experimental items.

Our main study is a connective insertion experiment involving 240 participants recruited via Prolific, with ideological diversity ensured through a pre-screening filter selecting self-identified liberals and conservatives in equal proportion. Participants are tested using two counterbalanced lists (120 participants per list); each list consists of 11 experimental items and 14 filler items, with the two lists together covering all 22 experimental items selected from the pre-test and sharing the same 14 fillers. Each item received 120 responses. Participants insert an explicit connective for each item. Political orientation was measured via (1) a 1–7 liberal–conservative self-placement Likert scale, (2) a 12-item Social and Economic Conservatism Scale [6], and (3) a three-item measure of political engagement. Connective insertion preferences will be analyzed using mixed-effects logistic regression models, with relation type as the dependent variable and different measures of political orientation as fixed effects, alongside random effects for participants and items. Political engagement and demographic variables will be included as covariates. Initial analyses show item-specific effects in the direction of the predicted ideological asymmetry. As a follow-up, we plan to replace the neutral speaker with one attributed to a specific political stance (liberal or conservative, e.g. “Republican” or “Democratic Party”), and hypothesize that participants can reason from the perspective of the specific speakers instead of prioritizing their own interpretation.

This work advances our understanding of discourse expectations in two ways. First, while prior work has linked belief systems to inferential biases at the lexical level [5] and in other pragmatic domains, we position political ideology as a systematic source of predictive bias specifically at the level of inter-sentential discourse relation interpretation, extending predictive processing accounts beyond linguistic experience and world knowledge. Second, it speaks to the nature and specificity of discourse expectations by showing how the same local linguistic material can support divergent global interpretations depending on the reader's belief system.

References

- [1] Eleni Miltsakaki, Rashmi Prasad, Aravind Joshi, and Bonnie Webber. 2004. The Penn Discourse Treebank. In *Proceedings of the fourth international conference on language resources and evaluation (LREC'04)*.
- [2] Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of the sixth international conference on language resources and evaluation (LREC'08)*.
- [3] Teun A. van Dijk. 1982. Opinions and attitudes in discourse comprehension. In *Advances in psychology*. Vol. 9. 35–51.
- [4] Merel Scholman and Vera Demberg. 2017. Crowdsourcing discourse interpretations: On the influence of context and the reliability of a connective insertion task. In *Proceedings of the 11th linguistic annotation workshop*. 24–33.
- [5] Niemi, L., Hartshorne, J., Gerstenberg, T., Stanley, M., & Young, L. (2020). Moral values reveal the causality implicit in verb meaning. *Cognitive Science*, 44(6), e12838.
- [6] Everett, J. A. (2013). The 12 item social and economic conservatism scale (SECS). *PloS one*, 8(12), e82131.