

**Commitment vs. discourse orientation : experimental and computational perspectives**  
 Grégoire Winterstein, Ghyslain Cantin-Savoie, Samuel Laperle, Josiane Van Dorpe and Nora Villeneuve

Département de Linguistique – Université du Québec à Montréal

In this work, we argue in favor of distinguishing between (i) the *commitment* associated with an utterance, i.e. the set of language-external situations compatible with the meaning of an utterance, and (ii) the *discourse orientation* of the utterance, i.e. the discourse possibilities made available by the utterance. This distinction is rooted in early observations by Anscombe and Ducrot (1983) about the differences between the informational and argumentative content of natural language expressions. We illustrate these differences with the adverbs *almost* and *barely*, as seen in (1).

1. a. John is done with his beer.
- b. John is *almost* done with his beer.
- c. John is *barely* done with his beer.

Intuitively, an utterance of (1b) entails that (1a) is false (Ducrot 1972, Jayez & Tovena 2008). Nevertheless, in many cases, substituting (1b) for (1a) will not affect the overall felicity of the discourse. For example, if someone asks who needs another beer from the bar, both (1a) and (1b) would intuitively convey that John might need one. This contrasts with (1c), which seems to commit its speaker to the truth of (1a), but would be understood as conveying that John does **not** need a beer. This is unexpected: if being *almost* done with one’s beer is grounds for ordering another, then being *completely* done (as is entailed by both (1b) and (1c)) should be even better grounds. But in the case of (1c), the opposite seems to be true. This conflict underlines the above distinction between *commitment* and *discourse orientation*. Jayez & Tovena (2008) account for these observations using a multilayered semantics in which *almost* conveys the negation of its prejacent via a conventional implicature (CI), and has the at-issue content in (2a) (adapted from J&T, where  $\text{std}(P)$  gives the standard degree of the property  $P$ ). Conversely, *barely* conveys the truth of its prejacent via a CI, and has the at-issue content in (2b).

2. a.  $[[\text{almost}]] = \lambda P.\lambda x. \text{deg}(P)(x) = d \ \& \ d > \text{std}(P) - \varepsilon$
- b.  $[[\text{barely}]] = \lambda P.\lambda x. \text{deg}(P)(x) = d \ \& \ d < \text{std}(P) + \varepsilon$

J&T argue that the discourse orientation of an utterance is determined solely by its at-issue content, ignoring CIs. Given the meanings in (2), *almost* picks out higher degrees than *barely*, which grounds their different discourse orientations.

In this work, we report the results of two experiments that verify the empirical validity of the distinctions reported above for analogous contrasts in French. Having established that naïve speakers do indeed distinguish between commitments and discourse orientation, we then discuss a third experiment showing the implications of that distinction for computational language models, from the perspective of textual entailment and natural language inference.

The first experiment was designed to test commitment. Participants ( $n=43$ ) were presented with out-of-context sentences. The target items instantiated one of the 3 conditions exemplified in (1):

(i) no modification (labeled  $\emptyset$ ), as in (1a), (ii) the use of *presque* ( $\approx$ ‘almost’) as in (1b), or (iii) *à peine* ( $\approx$ ‘barely’) as in (1c). Participants used a slider to situate the meaning of the sentence relative to two extrema. Extrema were chosen so that the  $\emptyset$ -condition would *a priori* denote the middle of the range, or some portion to its right. For example, in (1), the extrema were *30 minutes before John drinks the last drop of his beer* and *30 minutes after John drank the last drop of his beer*. Each participant saw 9 target items (3 in each condition), interspersed with 18 distractor items. Results are summarized in Fig. 1. We fitted linear mixed effect models with random intercepts for items and

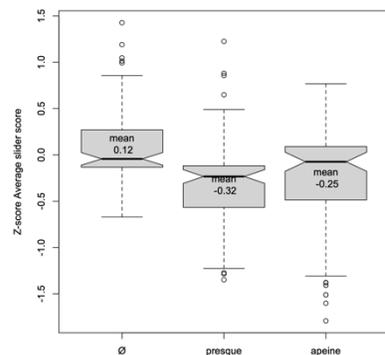


Fig. 1 - Experiment 1: Commitment

participants, and assessed the significance of our main factor via model comparison using likelihood ratio tests. We found a significant effect ( $\chi(1)=34.741$ ,  $p<0.001$ ), with the *presque* ( $\approx$ 'almost') condition being scored significantly below the other two. These results support the hypothesis that, with respect to commitment, the  $\emptyset$  and *à peine* ( $\approx$ 'barely') sentences correspond to comparable situations, distinct from those denoted by the *presque*-sentences, which are situated "below" the other two conditions.

The second experiment was designed to test discourse orientation. It involved the same conditions and material as in experiment 1, but sentences were presented after a context that *a priori* licenses the  $\emptyset$ -condition (as in the context described for (1) above). Participants ( $n=30$ ) judged the naturalness of the target sentence in the given context, using a 7-point Likert scale. Results are summarized in Fig. 2. Model comparison between ordinal mixed models with random intercept and slopes for items and participants shows a significant effect of the factor under study ( $\chi(1)=66.03$ ,  $p<0.001$ ), supporting the hypothesis that as far as discourse orientation is concerned, the  $\emptyset$  and *presque* sentences behave similarly, and are scored significantly higher than the *à peine* ones (in the full talk, we discuss contexts in which the relative acceptability of *à peine* and *presque* sentences is reversed). In summary, experiment 1 shows that with respect to commitment,  $\emptyset$  and *à peine* group together to the exclusion of *presque*, while experiment 2 shows that, with respect to discourse orientation,  $\emptyset$  and *presque* group together to the exclusion of *à peine*.

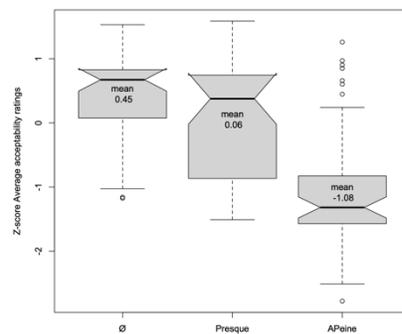


Fig. 2 - Experiment 2: Discourse orientation

These two experiments establish that speakers of French are indeed sensitive to the difference between commitment and discourse orientation. Our third experiment was designed to test whether state of the art language models such as BERT (Devlin et al., 2019) are similarly sensitive to this distinction. Distributional information plays a central role in shaping these models. For example, one of the training objectives of the BERT model is to predict whether, in a discourse of the form A B, the B sentence is a natural continuation of A (though this objective is not part of the training of all models, distributional information remains fundamental in their design Gastaldi, 2020). We therefore hypothesize that these models would be sensitive to the similarities exhibited in experiment 2 (*i.e.* discourse orientation), rather than to truth-conditional entailments like the ones of experiment 1 (*i.e.* commitment). To check this prediction, we tested inference patterns based on 200 semi-randomly selected examples from the French Wikipedia that involve the adverbs under study: *presque* and *à peine*. For each sentence, we used the pre-trained French CamemBERT model for natural language inference (Martin et al. 2020) to test whether the model predicts the truth of these sentences' prejacent. If the model is sensitive to commitment, it should predict the truth of the prejacent in the *à peine* cases and its negation for *presque*, and vice-versa if the model is sensitive to discourse orientation. The table on the right summarizes the average predicted probability of the prejacent and its negation for the *presque* and *à peine* cases. These results support the hypothesis that these models ground inference in discourse orientation rather than commitment.

	Infer prejac.	Infer ¬prejac.
<i>presque</i>	99.60	20.52
<i>à peine</i>	55.99	98.68

In the full paper, we discuss error patterns found in using BERT-like models for natural language inference tasks (*e.g.* Jiang & de Marneffe 2019), and show how these errors can be analyzed using the distinction between commitment and discourse orientation. We also argue that commitment and discourse orientation are confounded in the general case, which accounts for why this distinction has not yet been recognized in the computational literature.

**References:** Anscombe, JC & O. Ducrot *L'argumentation dans la langue* ♣ Devlin J. et al. (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding ♣ Ducrot O. (1972) Dire et ne pas dire ♣ Gastaldi, JL (2020) Why Can Computers Understand Natural Language? The Structuralist Image of Language Behind Word Embeddings. ♣ Jayez J. & L. Tovina (2008) Presque and almost: how argumentation derives from comparative meaning ♣ Jiang N. & MC de Marneffe (2019) Evaluating BERT for natural language inference: a case study on the CommitmentBank. ♣ Martin L. et al. (2020) CamemBERT: a Tasty French Language Model.