From local hesitations to global impressions of a speaker's feeling of knowing

Tanvi Dinkar
LTCI, Télécom Paris
IP Paris, France
tanvi.dinkar@telecom-paris.fr

Beatrice Biancardi
LTCI, Télécom Paris
IP Paris, France
beatrice.biancardi@telecom-paris.fr

Chloé Clavel
LTCI, Télécom Paris
IP Paris, France
chloe.clavel@telecom-paris.fr

1 Introduction

The aim of this work is to empirically study on a real-life dataset, whether the utterance level use of fillers can help in understanding/ interpreting the perception of the speaker that was formed by a listener. According to Brennan and Williams (1995), the listener's interpretation of the speaker's utterance includes estimates about the speaker's commitment to/ expressed confidence in what they are saying. Flavell (1979) termed these processes (of the speaker) as **metacognitive** ones, that is cognition about cognitive phenomena, or more simply "thinking about thinking". Research has linked fillers to the listener's assessment of a speaker's metacognitive state (Brennan and Williams, 1995). However, these results may not apply to spontaneous speech datasets collected in real-life contexts, or non-QA datasets. Additionally, the focus of analysis tends to be on utterances as if they occur in isolation, rather than part of an overall discourse. Thus existing studies do not focus on the connection between the hierarchical levels of discourse; i.e. how a speaker's local use of fillers could lead to a listener's overall (global) impression of the speaker. In this work, we study how does a speaker's use of fillers relate to the message from the speaker, and consequently, whether this relates to a listener's perception of the speaker. We do so by studying the POM dataset, a corpus of publicly available English monologue movie reviews (Park et al., 2014). Annotators (listeners) were asked to label the reviews for attributes such as "confidence"; without explicitly being told to pay attention to the speaker's use of fillers.

RQ1: (Local effect of fillers): How does a speaker's use of fillers relate to the message from the speaker? **H1**: Fillers are more likely to occur before the introduction of new and upcoming information in the review.

RQ2: (Global effect of fillers): How does the speaker's use of fillers relate to a listener's perception of the speaker? **H2**: the speaker's use of fillers preceding new information in the message contributes to the listener's perception of the speaker's confidence.

2 Methodology

To investigate **H1**, we consider the speaker's mention of entities related to the movie, that we extract from metadata files[1]. These entities could be categorised into actor, director or title of the movie. We inspect for each transcript, the distribution of filler positions, in relation to the automatically annotated entities in the discourse (denoted by Ent). We split these entities into Ent_new; i.e. entities newly introduced in the discourse, to indicate new information, and Ent_old to indicate entities already introduced in the discourse. Then, we check whether the distributions of filler positions (by its token position in the transcript) are significantly different compared to the distributions of 1. Ent_new and 2. Ent_old positions (by its first token's position), by utilising a Kruskal-Wallis H test with Benjamini-Hochberg correction. We then estimate the effect size by computing Cliff's Delta δ. Lastly, we compare the δ distributions of the two experiments, i.e. fillers with Ent_new versus fillers with Ent_old using a Wilcoxon signed-rank test.

To investigate **H2**, we take the mean of the three confidence labels provided by the three annotators. We then consider reviews that are categorised as high-confidence (HC; ratings ≥ 6, n=130) and low-confidence (LC; ratings ≤ 3, n=116). To calculate the percentage of fillers preceding new information (denoted by a new entity), we count the number of fillers in the review that occur before (but not

[1]The complete code and processed data will be made available online for reproducibility here https://github.com/tdinkar/fillers_in_POM.git

Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue, September 20–22, 2021, Potsdam / The Internet.

after) an Ent_new, constrained to a maximum distance of 1 token in between the filler and Ent_new. We normalise by dividing this count by the total number of fillers used in the review. From this, we obtain the percentage of fillers that occur before an Ent_new versus the percentage of fillers used in any other context that is not Ent_new. We then sum these two values for all HC and LC reviews, to get a cumulative percentage. We compute Odds Ratios (ORs) in order to investigate whether the use of fillers around new entities is associated with confidence, where the odds denote the outcome of HC or LC, given the occurrence of fillers before new entities, compared to the occurrence of fillers that do not occur before new entities. We expect that the more fillers are used in the context of preceding new entities, the greater the odds of HC.

3 Results and Discussion

H1: By Kruskal-Wallis H test the distributions of filler positions compared to 1. Ent_new and 2. Ent_old positions are significantly different for $\approx 15-20\%$ of the reviews (where $p \leq .05$). However, after utilising the Benjamini-Hochberg procedure for multiple testing correction, the distributions using this method do not significantly differ. This test is calculated using the sum of the ranks of each distribution. Given that the average review length is short ($\approx$ 256 tokens), and considering the close average median of fillers, Ent_new and Ent_old on reflection, this test may not capture nuances of the positional effects of fillers. However, by computing δ to estimate effect sizes, we found that for most reviews, fillers do occur before Ent_new (median = -0.30 , $SD = 0.41$), but not before Ent_old (median = 0.20, $SD = 0.37$), where the distributions of the δ values significantly differ ($Z = 27578.0$, $p < .05$ using Wilcoxon signed rank test). Majority of the reviews (565) have fillers occurring before Ent_new (sum of "large" to "small" δ sizes)[2], compared to 163 reviews that had negligible effect size, and 139 reviews that had positive effect size (reviews that had fillers occurring after the introduction of new entities). We see the opposite δ effect sizes for Ent_old, where most of the reviews have fillers occurring after entities already introduced in the discourse, but not before. Fillers occurring after Ent_old is entirely plausible given that new entities can occur throughout the review, and not just at the start of one. Given the large group with negligible effect size (247 reviews) for Ent_old, this does show that speakers may sometimes use fillers when repeating entities already introduced into the discourse. Compared to Dinkar et al. (2020), our findings suggest that there is more nuance to the way speakers utilise fillers (and indeed, our methodology is agnostic to sentence boundaries) in spontaneous speech. Therefore, regarding H1, **stylistically speakers do tend to use fillers when introducing a new entity rather than one already introduced** (whether intentionally or not remains an open question), and the positions of fillers w.r.t Ent_new significantly differ from positions of fillers w.r.t Ent_old.

H2: The results of the test show $OR = 0.72$ ($p < .001$, $95\%CI$ $0.6 - 0.8$). While $OR < 1$ (indicating that **the presence of fillers occurring before new entities gives a higher odds of LC), the presence of the stimulus on the outcome is small.** Interestingly, these findings are the opposite of what was expected (the speaker's use of fillers preceding new information contributes to the listener's perception of confidence; i.e. the more fillers are used in this way, the greater the odds of HC). According to the results, fillers occurring before new entities do not have a great effect on the odds of the HC (28% lower given the presence of new entities) rating that the listener gives the speaker. Inspecting the average rate of fillers in the review, it is clear that the use of fillers differs between HC and LC rated speakers (median filler rate of 0.026 and 0.045 respectively, with $U = 3873.0$ and $p < .05$ by Mann-Whitney U test). These results do not necessarily contradict Brennan and Williams (1995), i.e. there could be impressions formed by the listener about the speaker's expressed confidence based on fillers in spontaneous speech (as found in Dinkar et al. (2020)). However, these results would suggest that the effect may not be from fillers used in the context of introducing new entities. This is an interesting finding; as fillers in these contexts could still have a metacognitive function; i.e. the listener is drawn to the mind of the speaker and (their difficulties in) formulating a new referent as discussed in (Barr and Seyfeddinipur, 2010). But, it may be expected by the listener and and thus not necessarily contribute to the listener's perception of the speaker's expressed confidence.

[2]The magnitude of Cliff's Delta δ can be interpreted by using the thresholds from Romano et al. (2006), i.e. $|\delta| < 0.147$ "negligible", $|\delta| < 0.33$ "small", $|\delta| < 0.474$ "medium", and otherwise "large".

References

Dale J Barr and Mandana Seyfeddinipur. 2010. The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes*, 25(4):441–455.

Susan E Brennan and Maurice Williams. 1995. The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of memory and language*, 34(3):383–398.

Tanvi Dinkar, Pierre Colombo, Matthieu Labeau, and Chloé Clavel. 2020. The importance of fillers for text representations of speech transcripts. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7985–7993, Online. Association for Computational Linguistics.

John H Flavell. 1979. Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, 34(10):906.

Sunghyun Park, Han Suk Shim, Moitreya Chatterjee, Kenji Sagae, and Louis-Philippe Morency. 2014. Computational Analysis of Persuasiveness in Social Multimedia: A Novel Dataset and Multimodal Prediction Approach. In *Proceedings of the 16th International Conference on Multimodal Interaction*, ICMI 2014, page 50–57, New York, NY, USA. Association for Computing Machinery.

Jeanine Romano, Jeffrey D Kromrey, Jesse Coraggio, and Jeff Skowronek. 2006. Appropriate statistics for ordinal level data: Should we really be using t-test and cohen's d for evaluating group differences on the nsse and other surveys? In *annual meeting of the Florida Association of Institutional Research*, volume 177.