Filled Pauses in Writing: What can they Teach us about Speech?

Ralph ROSE <rose@waseda.jp>  Waseda University Faculty of Science and Engineering

Abstract

This exploratory study looks at filled pauses (uh, um) used in blog posts in order to see what light they may shed on pauses in spontaneous speech. Results show that writers use clause-boundary filled pauses in a relatively more selective manner, and thus use clause-internal filled pauses to highlight low frequency content words. When this written sample is compared with a spoken sample drawn from the Switchboard Corpus, the results suggest a hybrid model of filled pause use in speech combining a meaningful, word-like use of filled pauses with an automatic use driven by cognitive processes.

Background

Filled Pause Research

- FPs occur more commonly at clause boundaries than clause-externally (Rose 1996, Swerts 1998).
- FPs occur more commonly before content words than before function words (Maclay and Osgood 1959).
- FPs occur more commonly before low frequency words and before words with low contextual probability (Beattie and Butterworth 1979).
- Open FPs (uh) are more likely before shorter silent pauses while closed FPs (um) are more likely before longer silent pauses (Clark and Fox Tree 2002).

Based on such results, many researchers have concluded that speakers use FPs during lexical access when making choices among different alternatives (Schachter et al 1991) or to warn the listener of impending delay (Clark and Fox Tree 2002).

A continuing debate in the field is the lexical status of FPs. Some have argued that they are words (Clark and Fox Tree 1991) or to warn the listener of impending delay (Clark 1959, inter alia).

Discussion

Writers highlight low-frequency words with FPs. Results from the written corpus consistently show that blog writers see the dominant use of clause-boundary FPs as co-occurring with low-frequency content words. Writers seem to manipulate FPs in order to highlight low-frequency words. This is consistent with and supports the idea that FPs are used at times when lexical access is difficult (cf., Maclay and Osgood 1959, inter alia).

Writers highlight whole clauses with FPs. For clause boundary FPs, the written corpus suggests that FPs are used to no influence from the next word type, suggesting that the internal structure of the clause (or at least the first constituent) is not relevant to the use of FP. This contrasts sharply with the spoken corpus results.

Writers show little distinction between uh and um. While the spoken corpus reflects speaker preference for the open FP over the closed FP, writers do not show this preference. Thus, any distinction between the two is not salient to writers.

A hybrid model of FP use in speech. The difference between the spoken and written corpora results suggests a hybridized account of FPs: Some FPs are used in a conventionalized (i.e., word-like) manner to give a meta-comment on the upcoming utterance or some larger unit of discourse, or to draw special attention to the next word for some reason. Other FPs thus occur in an automatic manner—as a function of cognitive processes or strategies (e.g., repairs in Levitt 1983).

Future Work

- Explicitly test the hybrid model of FP use.
- Examine ways to identify the difference between conventionalized and automatic FPs in spontaneous speech.
- Examine the role that lexical density plays in the use of FPs in order to test word choice hypothesis.

References