Um, uh, and variation in American English

Matt Hunt Gardner
University of Oxford
matt.gardner@ling-phil.ox.ac.uk
&
Benedikt Szmrecsanyi
KU Leuven
benedikt.szmrecsanyi@kuleuven.be

Spontaneous speech takes place in real time, which means that speakers have to continually plan utterances both prior to and while speaking. Unsurprisingly, this is a demanding task. As a consequence, disfluencies are non-trivial in spontaneous speech; across studies about 6% of uttered words are disfluent [1][2]. Goldman-Eisler [3] shows that almost half of people's speaking time is made up of pausing and overt disfluencies like *um* and *uh*. There is overwhelming evidence from the psycholinguistic litterature that increased cognitive load coincides with increased rates of overt disfluencies (including filled pauses) and extended speech planning time (unfilled pauses). These two measures are in turn used as metrics for cognitive load. Exerting explicit executive control during production increases cognitive load, therefore choice-making is linked to disfluency and speech planning time. Grammatical planning involves both automatic processes and explicit control, and both must include probabilistic constraint-based variant selection. Variant substitution triggered by attention paid to speech or audience accommodation during the self-monitoring process is due to explicit executive control, adding to cognitive load, and potentially precipitating overt disfluencies or extended planning time. This leads to the hypothesis that grammatical alternations that are subject to style-shifting or audience design (i.e., sociolinguistic markers [4]) will coincide with a higher frequency of overt disfluencies and require more planning time compared to grammatical alternations that are not subject to style-shifting or audience design (i.e., sociolinguistic indicators [4]). Further, the greater the number of variable contexts subject to executive control within an utterance, the greater the number of overt disfluencies or the longer the amount of required planning time is expected.

No research has been conducted thus far to investigate whether specifically *grammatical* choice, i.e., the presence of grammatical variation, may cause increased disfluency in spontaneous speech. Against this backdrop we ask the question: does sociolinguistic variation make planning speech harder?

To answer this question we draw on the well-studied Switchboard Corpus of American English [5], which contains about 240 hours of recorded conversations between 542 Americans from all across the country. For each conversation we isolate variable contexts for 20 different morphosyntactic variables (N=46,867) and investigate correlations with both filled pauses (N=42,695) and speech planning time.

Surprisingly, we find that a greater number of variable contexts per 100 words coincides with fewer disfluent phenomena, suggesting variation instead facilitates speech production. This relationship is consistent across multiple varieties, and although factors like sex and age do have an effect on how many *um's* and *uh's* occur, the relationship between them and variable contexts does not. Our findings bolster arguments that variation is deeply embedded within the speech production process and is an integral component of the language faculty.

[1] Jean E. Fox Tree. "The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech". In: *Journal of Memory and Language* 34.6 (1995), pp. 709–738.

[2] Heather Bortfeld et al. “Disfluency Rates in Conversation. Effects of Age, Relationship, Topic, Role, and Gender”. In: *Language and Speech* 44.2 (2001), pp. 123–147.

[3] Frieda Goldman-Eisler. *Psycholinguistics. Experiments in Spontaneous Speech*. New York: Academic Press, 1968.

[4] William Labov. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press, 1972.

[5] J.J. Godfrey, E.C. Holliman, and J. McDaniel. “SWITCHBOARD: telephone speech corpus for research and development”. In: *[Proceedings] ICASSP-93. 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing* (1992).