MODEL ADAPTATION -- TO WHAT EXTENT DO SPEAKERS ADAPT WORD DURATIONS TO A DOMAIN?

Asad Sayeed (Saarland University), Stefan Fischer (Saarland University), Vera Demberg (Saarland University)

The Uniform Information Density (UID; e.g., Jaeger 2010) and related Smooth Signal Redundancy (Aylett and Turk, 2004) hypotheses imply that predictable words should have shorter spoken word durations than unpredictable words. This effect has been shown to hold in speech corpora using various language models as predictors (e.g., Jurafsky et al., 2001; Demberg et al., 2012; Seyfarth, 2014). First, we test whether semantic surprisal (which has not previously been used as a predictor) can predict spoken word durations over and above the effect of simpler n-gram and syntactic surprisal estimates; more interestingly, we examine the extent to which speakers adapt (see, e.g., Jaeger and Snider 2013) to the domain. Does the adaptation of word durations to semantic surprisal depend on the in-domain predictability of the words? In other words, do they immediately switch to an in-domain model, or is it governed by prior out-of-domain language experience? (Or anything in between? In case of adaptation, what mix of prior and posterior can we observe?)

Here we take a first step of answering these questions by training two extreme models, one that is trained entirely in-domain and a second one that is an out-of-domain model. To represent semantic similarities, we first construct a corpus-derived distributional vector space model whose dimensions are frequent content words. We then use these semantic similarities to reweight the probabilities from an n-gram model. The probability estimates of the resulting model are used to estimate semantic surprisal values.

Our study uses the AMI meeting corpus, which contains a record of 100 hours of meetings of up to five speakers and is annotated with orthographically correct transcriptions and exact word durations. Our in-domain model was trained on the same corpus using 10-fold cross validation, and the out-of-domain model was trained on a random sample from Gigaword. The in-domain language model has lower perplexity than the out-of-domain model and thus represents the objectively better language model for the corpus.

When evaluating the effect of estimated semantic surprisal on pronunciation times (in a linear mixed effects regression model also containing other predictors known to affect word durations, particularly word duration estimates from a TTS system, word frequency, simple ngram surprisal and syntactic surprisal), we found that the out-of-domain semantic surprisal estimates were a significant positive predictor of word durations for the whole range of semantic surprisal values, over and above the simpler predictors. On the other hand, the in-domain surprisal estimates were not positive predictors of word durations. More detailed analysis with spline models showed that the relationship between in-domain surprisal and residual word durations actually followed a u-shape. This indicates that the in-domain model under-estimates the surprisal (and hence word durations) of highly domain-predictable (but otherwise rare) words, and that speakers behaviour in this corpus was driven to a larger extent by their prior language model. Future work should investigate in more detail the time course of speaker adaptation in naturalistic communication situations.

References:

Aylett and Turk (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. JASA, 119(5):3048–3059.

Demberg, Sayeed, Gorinski, and Engonopoulos (2012). Syntactic surprisal has an effect on spoken word duration. EMNLP 2012.

Jaeger (2010). Redundancy and reduction: Speakers manage syntactic information density, Cognitive psychology 61 (1), 23-62.

Jaeger and Snider (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*,*127*(1).

Jurafsky, Bell, Gregory, & Raymond (2001). Probabilistic relations between words: Evidence from reduction in lexical production.*Typological studies in language*, *45*, 229-254.

Seyfarth (2014). Word informativity influences acoustic duration: effects of contextual predictability on lexical representation. Cognition, 133(1):140-155.