Informativity, predictability and partial pooling

Vsevolod Kapatsinski
University of Oregon
vkapatsi@uoregon.edu

Recent work in language production has observed that informativity predicts articulatory reduction of a linguistic unit above and beyond the unit's predictability in the local context, i.e., the unit's probability given the current context, e,g, [1-2]. Informativity of a unit is the inverse of average (log-scaled) predictability and corresponds to its information content. Research in the field has interpreted effects of informativity as speakers being sensitive to the information content of a unit in deciding how much effort to put into pronouncing it, or as accumulation of memories of pronunciation details in long-term memory representations. However, informativity can help improve an estimate of predictability simply because predictability estimates are noisy, especially when conditioned on rare contexts. Therefore, informativity can contribute to explaining variance in a dependent variable like reduction above and beyond local predictability simply because informativity improves the (inherently noisy) estimate of local predictability. This is the logic behind adaptive partial pooling in statistics: estimates of what would happen in a particular context should be based on observations of that context to the extent they are available, because they are the most relevant observations; however, when such relevant observations are few, observations of other similar contexts can help improve the estimate.

In the present study, predictability was defined as log transitional probability given the preceding word. Informativity is then defined as average susprisal across all tokens of the word. To illustrate the issue of noise in predictability estimates, suppose that the word *inched* occurs only once in the corpus. Then all words following it would have a probability of zero, except for the word that happened to occur after it (let's say *closer*), which would have a probability of 1. These estimates are highly unreliable and are unlikely to predict durations well, since they are based on a sample size of 1. By taking into account probabilities of words in other contexts, we might be able to estimate, for example, that *up* is a lot more probable than *higher* following *inched* and is likely to be pronounced more quickly.

We investigate two samples of words from the Switchboard Corpus [4] that are on the opposite ends of frequency, predictability, informativity and duration: 1) words that follow repetition disfluencies from [5], which are rare, unpredictable, informative and long, and 2) determiners in prepositional phrases, which are frequent, predictable, uninformative and short. Figure 1 shows that in both datasets, the predictability effect is strongest in frequent contexts, as also shown by [6], while the informativity effect is strongest in rare contexts. This is what we would expect if informativity helped improve predictability estimates, as the informativity effect is strongest where predictability is least reliable.

We then investigate how large the informativity effect and its interactions with context frequency would be if it were due entirely to noise in predictability estimates through simulations that assume the same bivariate relationship between predictability and duration seen in the real data, and the same random effect of word, but no real effect of informativity. These simulations suggest that not all of the informativity effect is due to noise in predictability. This result suggests that speakers themselves use adaptive partial pooling in integrating word predictability or duration estimates across contexts with those observed in the current context. Adaptive partial pooling in language production can be implemented both by mechanisms that perform adaptive partial pooling explicitly (i.e., hierarchical regression), and those that do so implicitly (large language models or exemlar models with differential exemplar activation based on distance from the probe).

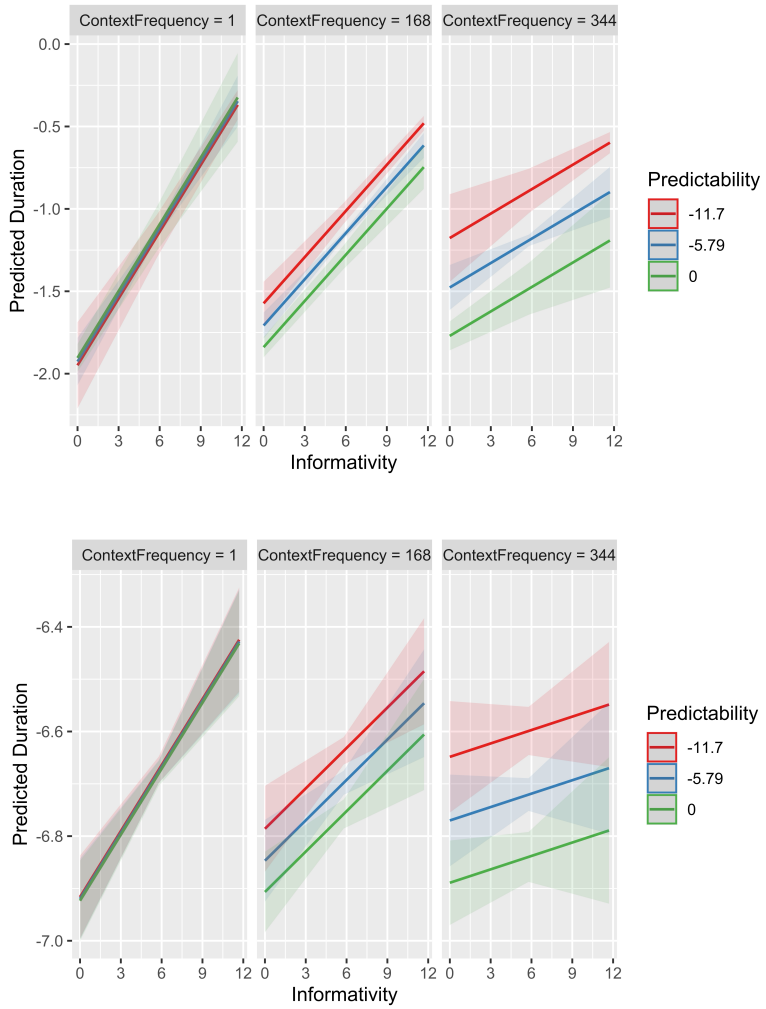

Figure 1. Interactions of Informativity (x axis) and Predictability (lines) with the square root of context frequency (panel) from a mixed-effects model. Top row: Post-disfluency words. Bottom row: Determiners in prepositional phrases. The three lines and three panels show the effect for the minimum, mean and maximum (in the dataset) value of each predictor in the Post-disfluency dataset. Predictability has no effect in the rarest contexts, while informativity has only a weak effect in the most frequent contexts.

References

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

[2] Sóskuthy, M., & Hay, J. (2017). Changing word usage predicts changing word durations in New Zealand English. Cognition, 166, 298-313.

[3] Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. In Proceedings of the International Conference on Acoustics, Speech, & Signal Processing (ICASSP) '92, March 23-26, San Francisco, CA (pp. 517–520).

[4] Harmon, Z., & Kapatsinski, V. (2021). A theory of repetition and retrieval in language production. Psychological Review, 128(6), 1112-1144.

[5] Jaeger, T. F. (2006). Redundancy and syntactic reduction in spontaneous speech (Doctoral dissertation, Stanford University).