



# N-gram frequency effects on speech production in German and Mandarin Chinese

Ching Chu Sun & Peter Hendrix

## Elicited speech

Two naming experiments were conducted in German and in Mandarin Chinese. The stimuli in both experiments were random three-word-sequences. We transformed trigram frequency counts as well as component word bigram and unigram frequencies using principal component analysis with varimax rotation to resolve collinearity in the data. We controlled for the effects of other lexical-distributional variables that are known to influence pronunciation durations.

We analyzed the experimental data using a quantile generalized additive mixed model (QGAMM). In addition to effects of control variables, the QGAMM analysis revealed effects of unigram and bigram frequency. Crucially, the effect of trigram frequency was highly significant in both languages as well. Consistent with the results of previous studies, pronunciation durations are shorter for word trigrams with a higher frequency in German and Mandarin Chinese.

## Spontaneous speech

We furthermore investigated the effect of trigram frequency in spontaneous speech. Pronunciation durations of random three-word-sequences were extracted from a conversational corpus of German as well as a spoken corpus of Taiwan Mandarin. Word trigrams that were preceded or followed by pauses were excluded from the data. As before, we applied a principal component analysis with varimax rotation to ensure that the frequencies of unigrams, bigrams and trigrams were orthogonal.

A QGAMM analysis of the data revealed significant effects of trigram frequency over and above the effects of the component unigram and bigram frequencies in both German and Mandarin Chinese. The effects of trigram frequency effects were stronger in spontaneous speech as compared to elicited speech. We conclude that the frequency of multiword sequences influences pronunciation durations, even when these multiword sequences do not form fixed expressions of lexical bundles.

