Effects of Position, Stress and Manner of Articulation on Consonant-Vowel Co-occurrence in Three Languages

Eleonora Albano
University of Campinas

This is a follow-up on a previous study (Albano, 2016) of the effects of position, stress and manner of articulation on consonant-vowel (CV) co-occurrence frequencies in large lexical databases of British English (BE) and Brazilian Portuguese (BP). The present study looks into the same effects in the token frequencies of the original databases with the addition of a new one, from Latin American Spanish (LAS). The aim is to inquire whether such effects, very reliably attested in the three lexicons, are also present in language use.

The original findings for type frequency were: (1) lingual consonants (coronals and dorsals) present biomechanically driven CV biases in the two lexicons; (2) such biases have an uneven distribution, being especially strong in syllables beginning with stops and affricates in initial unstressed position; (3) log-linear modeling reveals significant interactions between the factors initial, unstressed, obstruent, lingual, and front or back. The main follow-up question is whether the conditions favoring such biases take part in word selection in lexicons. If they do, token frequencies should display the same biases, and should correlate highly with type frequencies.

To probe into such issues, we had to avoid four methodological pitfalls: (1) the existence of morphological confounds in the data; (2) the low power of non-parametrical statistics; (3) the need for multiple comparisons; and (4) the confounding effect of over-high token frequencies. These were handled through a mixture of conservative and innovative techniques. Conservativeness was necessary to assure comparability with the previous literature. Innovativeness was necessary to assure reliability, given the inability of the traditional approaches to handle multiple factors.

The three databases were CELEX, for BE, Call Home, for LAS, and LAEL, for BP. CELEX and Call Home are available from Linguistic Data Consortium; LAEL is available from the applied linguistics program of the Pontifical Catholic University of São Paulo. All are large enough (over 40,000 words). CELEX and Call Home come with phonemic transcription, while LAEL is easy to convert from orthography. In line with tradition, consonants were classified as labial, coronal, and dorsal; and vowels, as front, central, and back. In addition, CV pairs were coded as to position (initial and medial), stress (stressed and unstressed), and C manner of articulation (stops/affricates, fricatives and sonorants). Statistical analyses included simple and nested crosstabs and log-linear modeling.

The words responsible for the morphological biases were filtered from the original databases, preserving around 90% of their contents. Token frequencies above 10,000 were also filtered. In addition, Pearson’s correlation coefficients were calculated for log-normalized type and token frequencies of all the CV pairs in the three databases. In crosstab analyses, Cramér’s V and adjusted residuals supplemented chi-squared to assess association strength and cell significance. In log-linear analyses, a reduced taxonomy helped cut down the number of factors. Thus, coronals and dorsals were collapsed into linguals, and stops/affricates and fricatives were collapsed into obstruents. Both techniques generate a number of p values requiring a multiple comparison procedure. We adopted Benjamini and Hochberg’s (1995), which controls for both type I and type II error through a quantity called false discovery rate, which maintains both α and β below cutoff (set at 5%). The use of these statistics with Call-Home’s type frequencies confirmed the previous findings; their application to the three databases’ token frequencies suggested the existence of similar trends in language use.

The hypotheses were: (1) CV types and tokens correlate highly; (2) tokens reproduce the biomechanically driven patterns found in types; and (3) tokens exhibit significant interactions between the factors initial, unstressed, obstruent, lingual, and front or back.

The results corroborate the three hypotheses. Figure 1 shows the dispersion diagram for the log-normalized type and token frequencies of the CV pairs in CELEX. Figure 2 exhibits the residuals for the type frequencies in initial unstressed position in Call Home. Figure 3 exhibits the residuals for the
token frequencies in initial unstressed position in LAEL. Space restricts each illustration to one language, but patterns are recurrent across all others.

Figure 1 – Dispersion for CELEX’s log type and token frequencies; R = 0.88 p <0.0001.

Figure 2 – Adjusted residuals for CALL HOME type frequencies with initial unstressed stops.

Figure 3 - Adjusted residuals for LAEL token frequencies with initial unstressed stops.

High correlations show that the frequency trends of language use resemble those of lexicons. In addition, the biases are the same across counts and languages: while labials have fewer biases, coronals seek front vowels and dorsals avoid them. Such patterns are likely to arise from an interplay between biomechanical and speech planning constraints. First, the preferred CV combinations are among those that Nam et al.’s simulation (2013) showed to be the most synergetic. Second, positive and negative biases are predictable from the degree of articulatory constraint model of co-articulation and Articulatory Phonology (Recasens, Pallarès, & Fontdevila, 1997; Iskarous et al., 2013). Third, the interactions between the factors initial, unstressed, obstruent, lingual, and front or back are always significant. Fourth, initial unstressed position always displays the highest association strengths.

Why should initial unstressed position enhance biomechanically driven CV preferences? Clues in the laboratory phonology literature suggest that this effect originates in the complexity of initiation (MacNeilage & Davis, 2005; Mooshammer et al., 2012). Since initial strengthening (Keating, Cho, Fougeron, & Hsu, 2003) and stress contour buildup (De Jong, Beckman, & Edwards, 1993) place their own demands on co-articulation, the preference for synergetic CVs might be a means of facilitating gestural timing in initial unstressed position.

Overall, the results suggest that biomechanical and speech-planning constraints interact in language use, eventually making their way to lexical phonotactics. However, since the languages involved are interrelated, further research is necessary to bolster this interpretation.

References


