

The Gradient Phonotactics of English CVC Syllables

Olga Dmitrieva & Arto Anttila

Department of Linguistics, Stanford University, USA; dmitro@stanford.edu

This study examines the factors affecting the gradient well-formedness of English CVC syllables. We will focus on two major effects: place dissimilation between the two consonants and the prominence alignment between stress, vowel height, and consonant place. First, recent research has shown that the Obligatory Contour Principle (OCP) emerges gradually in phonotactics. In particular, Berkley (1994) demonstrated a gradient OCP-place effect in the English lexicon: monomorphemic monosyllabic words where the onset and the coda share place of articulation are statistically underrepresented. We replicate and extend Berkley's result using CELEX2 (Baayen, Piepenbrock, & Gulikers, 1995) and CMU (Weide, 1998) dictionaries, showing that the gradient OCP-place effect emerges not only in monomorphemic monosyllabic words, but in all CVC syllables of English. Second, we point out a gradient prominence alignment effect among stress, vowel height, and consonant place, which requires that the most prominent element in each domain combines with the most prominent elements in other domains.

In our study, we code each CVC syllable in the CELEX and CMU databases for (1) the place feature of the onset and the coda (dorsal, labial, and coronal), (2) vowel height/sonority (low and mid vs. high and reduced), and (3) stress of the syllable (stressed and unstressed). The degree of the representation of each syllable type in the lexicon is quantified in terms of the Observed/Expected frequency ratio (Frisch, Pierrehumbert, & Broe, 2004). Expected frequency was calculated by multiplying the probabilities of each term, based on the assumption that they are independent. Resulting O/E values show that the CVC syllables with the homorganic consonants in onsets and codas are underrepresented in the dataset compared to the syllables with the heterorganic onsets and codas. We also observe that low and mid vowels are underrepresented in combinations with coronal consonants and in unstressed syllables, while high and reduced vowels are underrepresented in combinations with labial and dorsal consonants and in stressed syllables.

To evaluate the size of the effect, we apply a multiple regression analysis (Coetzee & Pater, 2008). The prohibition against low/mid vowels in unstressed syllables and high/reduced vowels in stressed syllables emerges as the most robust factor ($p < 0.001$). The OCP effect emerges as the second most predictive factor ($p < 0.001$). Finally, the prohibition against marked consonants (dorsal, labial) in syllables that were unstressed or had low-sonority vowels (high, reduced) also reached significance ($p < 0.01$). This effect is more pronounced for onsets than for codas. Interestingly, the symmetric counterpart of the last factor, that is, the prohibition against unmarked consonants (coronal) in syllables that were stressed or had high-sonority vowels (low, mid) vowels was not significant. These results suggest that vowels in English CVC syllables undergo gradient neutralization in non-prominent positions and augmentation in prominent positions, while consonants undergo neutralization alone. Thus, stressed syllables achieve both acoustic salience (through more sonorous vowels) and informational efficiency (through an extended consonantal inventory). Unstressed syllables, on the other hand, reduce the vowel and consonant inventory to the least prominent segments.

The study suggests a number of further questions. The most important one is the possible effect of the inflectional and derivational morphology on the statistical patterns in the data. In particular, phonological and prosodic characteristics of inflectional endings and derivational affixes could be responsible for boosting certain quantitative patterns, such as overrepresentation of coronal consonants in the unstressed syllables. A preliminary investigation using the CELEX corpus involved excluding all final syllables that are likely to carry most of the morphological suffixes. Excluding morphological suffixes did not appear to cause significant changes in the observed patterns. This suggests that the presence of morphologically complex items in the dataset only serves to magnify the prominence alignment effects, but is not the source of these effects. A more detailed investigation of the morphological factors is left for the future work.

The study has one particularly interesting theoretical consequence. The relationship between vowel sonority and syllable prominence is well attested, and phonetically grounded. Stress is often attracted to

acoustically more prominent (longer, louder) low and mid vowels. Vowels can undergo augmentation in stressed syllables and reduction in unstressed syllables. The connection between consonant place and prominence, on the other hand, is controversial. De Lacy (2001) argues that constraints relating prominence and place of articulation, e.g. *STRESS/LABIAL, cannot exist since no phonological alternations of this kind are attested. Our study challenges this assumption by showing that there is a connection between the place feature of the consonants and syllable prominence. We find a statistically significant preference for coronal onsets and codas over dorsal and labial onsets and codas in unstressed CVC syllables.

There is also some evidence that the connection between consonant place and prominence may be phonetically motivated. Compared to dorsals and labials, coronals require less articulatory effort and time. At least for coronal stops, in some cases this leads to less overlap with neighboring sounds and lower perceptibility. In contrast, dorsal and labial consonants involve larger and slower moving articulators, which results in more coarticulation in the adjacent sounds and enhanced perceptual salience (Jun, 2004).

We propose an optimality-theoretic account for the data, which models the gradient well-formedness of syllables in terms of their relative grammatical complexity measured in terms of ranking information (Anttila, 2008). Grammatical complexity is a direct consequence of the relative markedness of the syllable: the more ranking information is required for the syllable to surface faithfully, the more grammatically complex it is. The model predicts that the grammatically most complex syllables should be least well-formed and least frequent, and the less grammatically complex the syllable is, the more well-formed and more frequent it should be. We show that the model correctly predicts a number of quantitative asymmetries in the data. The grammar benefits significantly from the results of the statistical analysis: the constraints were chosen so as to correspond to the factors that reached significance in the regression analysis, e.g. *X/i ‘Avoid a combination of stressed syllable and high vowel’ and the rankings among the constraints were introduced based on their factors’ coefficients in the regression analysis. As a result, the advantages of the statistical techniques combined with the devices of the formal grammar make the proposed approach adequate from both a descriptive and explanatory standpoints.

References

- Anttila, A. (2008). Gradient phonotactics and the Complexity Hypothesis. To appear in *Natural Language and Linguistic Theory*.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database* (Release 2). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].
- Berkley, D. (1994). The OCP and gradient data. *Studies in the Linguistic Sciences*, 24, 59–72.
- Coetzee, A., & Pater, J. (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. To appear in *Natural Language and Linguistic Theory*.
- de Lacy, P. (2001). Markedness in prominent positions. In O. Matushansky, A. Costa, J. Martin-Gonzalez, L. Nathan, & A. Szczegielniak (Eds.), *HUMIT 2000: MITWPL 40* (pp. 53–66). Cambridge, MA: MIT Working Papers in Linguistics.
- Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory*, 22(1), 179–228.
- Jun, J. (2004) Place assimilation. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically-based Phonology* (pp. 35–57). Cambridge: Cambridge University Press.
- Weide, R. (1998). *The CMU pronunciation dictionary* (Release 0.6). Carnegie Mellon University. Available online at <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>.