# The Gradient Phonotactics of English CVC Syllables

**Olga Dmitrieva & Arto Anttila**
**Department of Linguistics, Stanford University**

## Introduction

Factors affecting the well-formedness of English CVC syllables:

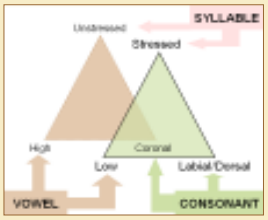
- *The OCP-place*: a gradient prohibition against homorganic consonants in the onset and the coda of the CVC syllables.

*E.g. gag vs. gap*

HYPOTHESIS: Syllables with C1 and C2 of the same place of articulation are underrepresented

- *The prominence alignment between syllable stress, vowel height, and consonant place:*

HYPOTHESIS: Syllables that violate prominence alignment are underrepresented




## Methods

**Material:**

- CMU pronunciation dictionary and CELEX lemma lexicon
- Only CVC syllables with primary or no stress
- Syllables - stressed and unstressed
- Consonants - coronals, dorsals, and labials
- Vowels - high (high + reduced) and low (low + mid)

Repeat → [ɪ], [pit]

*25,888 CVC syllables from CELEX*
*83,798 CVC syllables from CMU*

**Effect size evaluation:**

- Observed frequency/Expected frequency ratio (O/E ratio):

*P(dorsal-V-dorsal) = P(onset=dorsal) * P(coda=dorsal)*
*E(dorsal-V-dorsal) = P(dorsal-V-dorsal) * Total*

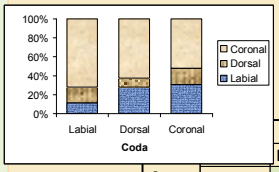
O/E ratio > 1.00 overrepresentation
O/E ratio < 1.00 underrepresentatio

- Multiple regression

## Results

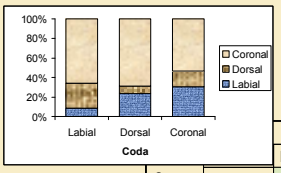
**1. Syllables that violate OCP-place are underrepresented:**

Onset-coda cooccurrences (O/E values):




CELEX

| | | Coda | | |
|---|---|---|---|---|
| | | Labial | Dorsal | Coronal |
| Onset | Labial | 0.40 | 0.77 | 1.23 |
| | Dorsal | 0.78 | 0.42 | 1.25 |
| | Coronal | 1.29 | 1.20 | 0.86 |



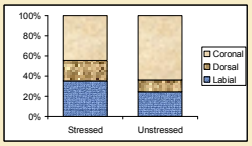

CMU

| | | Coda | | |
|---|---|---|---|---|
| | | Labial | Dorsal | Coronal |
| Onset | Labial | 0.31 | 0.84 | 1.09 |
| | Dorsal | 1.63 | 0.51 | 1.03 |
| | Coronal | 1.17 | 1.21 | 0.94 |

**2. Syllables that violate consonant-stress alignment are underrepresented:**

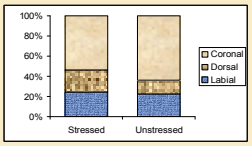
a. Labial and dorsal consonants in unstressed syllables
b. Coronal consonants in stressed syllables

CMU




| | Onset | | Coda | |
|---|---|---|---|---|
| | Stressed | Unstressed | Stressed | Unstressed |
| Labial | 1.27 | 0.87 | 1.68 | 0.65 |
| Dorsal | 1.30 | 0.79 | 0.96 | 1.03 |
| Coronal | 0.80 | 1.14 | 0.95 | 1.04 |

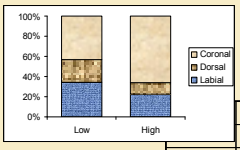
CELEX




| | Onset | | Coda | |
|---|---|---|---|---|
| | Stressed | Unstressed | Stressed | Unstressed |
| Labial | 1.15 | 0.83 | 1.18 | 0.88 |
| Dorsal | 1.21 | 0.86 | 0.99 | 1.01 |
| Coronal | 0.82 | 1.12 | 0.88 | 1.03 |

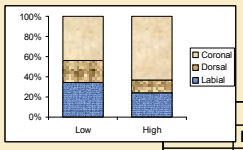
**3. Syllables violating consonant-vowel alignment are underrepresented:**

a. Low vowels with coronal consonants
b. High vowels with labial or dorsal consonants




CELEX

| | Onset | | Coda | |
|---|---|---|---|---|
| | Low | High | Low | High |
| Labial | 1.28 | 0.83 | 1.12 | 0.93 |
| Dorsal | 1.41 | 0.64 | 1.03 | 0.87 |
| Coronal | 0.76 | 1.15 | 0.96 | 1.03 |

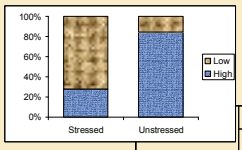



CMU

| | Onset | | Coda | |
|---|---|---|---|---|
| | Low | High | Low | High |
| Labial | 1.23 | 0.94 | 1.87 | 0.68 |
| Dorsal | 1.42 | 0.70 | 1.06 | 0.90 |
| Coronal | 0.78 | 1.11 | 0.92 | 1.04 |

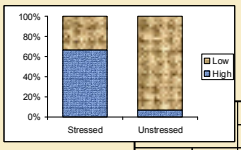
**4. Syllables that violate vowel-stress assignment are underrepresented:**

a. Low vowels in unstressed syllables
b. High vowels in stressed syllables




CELEX

| | | Syllable type | |
|---|---|---|---|
| | | Stressed | Unstressed |
| Vowel quality | Low | 1.89 | 0.41 |
| | High | 0.45 | 1.37 |




CMU

| | | Syllable type | |
|---|---|---|---|
| | | Stressed | Unstressed |
| Vowel quality | Low | 2.42 | 0.24 |
| | High | 0.46 | 1.29 |

## Regression

Cases: 36 syllable types:
3 onset place * 3 coda place * 2 stress * 2 vowel height
e.g. *LLHS* - labial-labial, high vowel, stressed

**CMU**

| Factors | Coefficient | Std. Error | t | p |
|---|---|---|---|---|
| Expected | 0.857 | 0.077 | 11.120 | <0.001 |
| OCP | -0.961 | 0.204 | -4.709 | <0.001 |
| Stressed/High | -1.708 | 0.254 | -6.723 | <0.001 |
| Unstressed/Low | -1.169 | 0.257 | -4.557 | <0.001 |
| Lab&Dor onset/Unstressed | -0.820 | 0.251 | -3.268 | <0.01 |
| Lab&Dor coda/Unstressed | -0.737 | 0.250 | -2.947 | <0.01 |

R = 0.945 ($F(6, 35) = 13.515$, $p < 0.001$)

**CELEX**

| Factors | Coefficient | Std. Error | t | p |
|---|---|---|---|---|
| Expected | 0.856 | 0.088 | 9.689 | <0.001 |
| OCP | -0.783 | 0.16 | -4.599 | <0.001 |
| Stressed/High | -1.830 | 0.2 | -4.151 | <0.001 |
| Unstressed/Low | -1.513 | 0.199 | -7.591 | <0.001 |
| Lab&Dor onset/High | -0.660 | 0.196 | -3.372 | <0.01 |
| Lab&Dor coda/High | -0.477 | 0.196 | -2.43 | <0.05 |

R = 0.943 ($F(6, 35) = 38.689$, $p < 0.001$)

Dependent variable:
- Log of the observed frequency

Independent variable:
- Log of the expected frequency
- Binary coded phonotactics factors:
  *1 – violates, 0 – does not violate*

**In CMU significant factors:**
Vowel-stress alignment
OCP
No labial/dorsal in unstressed syllables

**In CELEX significant factors:**
Vowel-stress alignment
OCP
No labial/dorsal with high vowels

## OT Analysis

**Constraints (significant regression factors):**

**OCP** No homogeneous consonants in onset and coda of the same syllable
***x/a** Low vowels are prohibited in unstressed syllables
***X/I** High vowels are prohibited in stressed syllables
***x/p_** No labial/dorsal onsets in unstressed syllables
***x/_p** No labial/dorsal codas in unstressed syllables
***p_/I** No labial/dorsal onsets followed by high vowels
***i/_p/** No labial/dorsal codas preceded by high vowels
**Faith**

- Markedness constraints ranked according to regression coefficients
- Markedness constraints unranked with respect to Faith

Multiple grammars
↓
Implicational hierarchies (T-orders)
↓
Markedness into frequency

Measure of the model fit:

| Data | Precision | Recall |
|---|---|---|
| CMU | 0.75 | 0.88 |
| CELEX | 0.74 | 0.87 |

## Conclusions

- The gradient OCP-place effect is active in all CVC syllables of English (not only in monosyllabic words)
- Prominence alignment in CVC syllables:
  Best stressed syllable – with low or mid vowels
  Best unstressed syllable – with high or reduced vowels and coronal consonants
- Positional neutralization and augmentation for vowels
- Only positional neutralization for consonants

**References:**
Anttila, A. (2008). Gradient phonotactics and the Complexity Hypothesis. To appear in *Natural Language and Linguistic Theory*.
Anttila, A. & Curtis, A. (2006). T-Orders. Stanford University, ms.
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*.
*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.*

**Laboratory Phonology 11, 30 June - 2 July 2008, Wellington, New Zealand**