

# GRADIENT OCP AND HARMONIC ALIGNMENT IN ENGLISH PHONOTACTICS\*

Olga Dmitrieva, Matthew Adams, Jason Grafmiller,  
Scott Grimm, Yuan Zhao, Arto Anttila

The 82nd Annual Meeting of the Linguistic Society of America  
3 January 2008

## 1. Introduction

- (1) The Obligatory Contour Principle (OCP) prohibits adjacent identical elements (Leben 1973, Goldsmith 1976, McCarthy 1986, Odden 1986, Yip 1988, Coetzee 2004, among others).
- (2) Berkley (1994) demonstrated gradient place and manner OCP effects between English consonant pairs within monomorphemic monosyllables. The OCP effect weakens with the amount of intervening material between the consonants: words of type *bob*, *bomb* show a stronger OCP effect than words of type *babe*.
- (3) Our goals:
  - (a) Replicate Berkley's results.
  - (b) Explore the OCP effect across syllables that differ in prominence, in particular:
    - stressed vs. unstressed syllables
    - syllables with high vs. low vowels.
- (4) Our finding:
  - (a) The gradient Place-OCP effect in English is active not only in monomorphemic monosyllabic words, but in all CVC syllables.
  - (b) Harmonic alignment between consonant place, vowel sonority, and syllable prominence in English:

Asymmetry 1: stress and vowel sonority

- Stressed syllables prefer non-high and non-central vowels
- Unstressed syllables prefer high and reduced vowels.

Asymmetry 2: stress and consonant place

- Stressed syllables prefer labial and dorsal (marked) onsets and codas.
- Unstressed syllables prefer coronal (unmarked) onsets and codas.
- In both cases this preference is stronger for onsets than for codas.

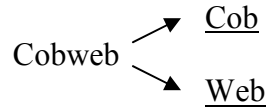
---

\* We would like to thank Uriel Cohen Priva and Giulio Caviglia for scripting help, Jeremy Glick for ongoing help with statistical analysis, and the audience of Stanford Phonology workshop for helpful comments and suggestions.

## 2. OCP in words – OCP in syllables

(5) Procedure:

(a) All syllables in the CMU pronunciation dictionary and CELEX lemma lexicon were considered as separate tokens:



(b) Only CVC syllables.

(c) Only syllables with primary stress and no stress.

(d) Only syllables with simple vowels (no diphthongs or nasalized vowels).

(e) Resulting data files:

- 84,346 CVC syllables from CMU database
- 25,888 CVC syllables from CELEX database

(6) All onsets and codas were classified into coronal [t, d, s, z, ʃ, ʒ, tʃ, dʒ, r, l, n, j], dorsal [k, g, h, ŋ], and labial [p, b, f, v, m, w].

(7) The strength of the place-OCP effect for onset-coda pairs in CVC syllables was calculated in terms of Observed/Expected (O/E) values (Frisch et al. 2004):

- O = observed frequency of cooccurring onset-coda pairs.
- E = expected frequency of cooccurring onset-coda pairs. Calculated with the assumption that onset and coda place are independent events:  
 $P(\text{dorsal-V-dorsal}) = P(\text{onset=dorsal}) * P(\text{coda=dorsal})$   
 $E(\text{dorsal-V-dorsal}) = P(\text{dorsal-V-dorsal}) * \text{Total}$
- An O/E value greater than 1 indicates that there are more observed combinations than expected, i.e. the combination is favored.
- An O/E value smaller than 1 indicates that there are fewer observed combinations than expected, i.e. the combination is disfavored.

(8) Table 1: O/E values for onset-coda cooccurrences in CVC syllables of CMU.

Onset	Coda			
		Dorsals	Labials	Coronals
	Dorsals	<b>0.50*</b>	1.67*	1.03*
	Labials	0.84*	<b>0.31*</b>	1.09*
	Coronals	1.21*	1.16*	<b>0.95*</b>

\* - Value significantly different from 1.00 by chi-square.

(9) Table 2: O/E values for onset-coda cooccurrences in CVC syllables of CELEX.

Onset	Coda			
		Dorsals	Labials	Coronals
	Dorsals	<b>0.60*</b>	1.08	1.11*
	Labials	1.04	<b>0.42*</b>	1.13*
	Coronals	1.08*	1.25*	<b>0.91*</b>

\* - Value significantly different from 1.00 by chi-square.



#### 4. Vowel sonority and stress

(17) Observations: Stress is often attracted to the syllables with the most sonorous vowels; short and reduced vowels often cannot be stress-bearing; vowels in unstressed syllables are often reduced (see e.g. Kenstowicz 1994, de Lacy 2001).

(18) Procedure: We classified all the vowels in the corpus into LOW vowels, which include all non-high vowels, and HIGH vowels, which include all high vowels and schwa.

(19) Table 6: Vowel quality and syllable prominence in CMU (O/E values).

		Syllable type	
		Stressed	Unstressed
Vowel quality	Low vowels	<b>2.42*</b>	0.24*
	High vowels and schwa	0.46*	<b>1.29*</b>

\* - Value significantly different from 1.00 by chi-square.

(20) Table 7: Vowel quality and syllable prominence in CELEX (O/E values).

		Syllable type	
		Stressed	Unstressed
Vowel quality	Low vowels	<b>1.89*</b>	0.41*
	High vowels and schwa	0.45*	<b>1.37*</b>

\* - Value significantly different from 1.00 by chi-square.

(21) As expected, low vowels are over-represented in stressed syllables, while high and reduced vowels are over-represented in unstressed syllables.

(22) Chi-square for the relationship between the two factors (vowel quality and stress) was significant (effect size of **0.640** in CMU and **0.572** in CELEX, measured by Phi and Cramer's V):

(a) In CMU:  $\chi^2(1, N = 84,346) = 34558.82, p < .001$ .

(b) In CELEX:  $\chi^2(1, N = 25,888) = 8483.39, p < .001$ .

Table 8: Vowel Height \* Stress crosstabulation (CMU).

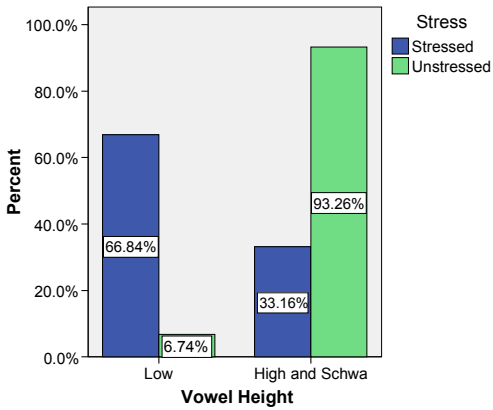
		Stress		Total
		Stressed	Unstressed	Stressed
Vowel Height	Low	19556	3711	23267
	High	9701	51378	61079
Total		29257	55089	84346

(23) Constraints:

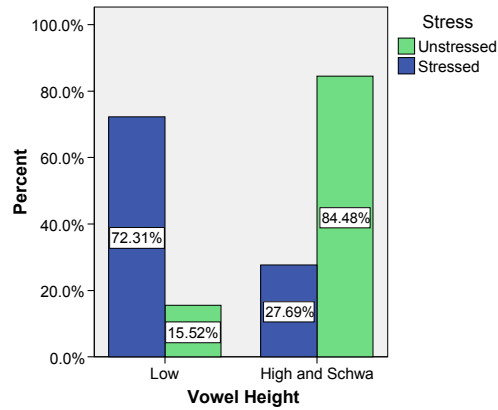
\*X/I            Avoid stressed syllables with high vowels.

\*x/A            Avoid unstressed syllables with low vowels.

Graph 1: Vowel quality and stress in CMU



Graph 2: Vowel quality and stress in CELEX



### 5. Consonant place and stress

(24) Further observations:

- (a) All labial/dorsal initial stressed syllables have *higher* O/E values than labial/dorsal initial unstressed syllables.
- (b) All coronal initial stressed syllables have *lower* O/E values than coronal initial unstressed syllables.

(25) Table 9: O/E values for onset-coda cooccurrences in stressed and unstressed CVC syllables of CMU.

		Coda					
Onset		Dorsals		Labials		Coronals	
		Unstressed	Stressed	Unstressed	Stressed	Unstressed	Stressed
	<b>Dorsals</b>	<b>0.36</b> ←	<b>0.75</b>	0.94 ←	3.02	0.88 ←	1.30
	<b>Labials</b>	0.73 ←	1.05	<b>0.07</b> ←	<b>0.78</b>	0.96 ←	1.34
	<b>Coronals</b>	1.34 →	0.96	0.85 ←?	1.75	<b>1.10</b> →	<b>0.66</b>

(26) Table 10: O/E values for onset-coda cooccurrences in stressed and unstressed CVC syllables of CELEX.

		Coda					
Onset		Dorsals		Labials		Coronals	
		Unstressed	Stressed	Unstressed	Stressed	Unstressed	Stressed
	<b>Dorsals</b>	0.58 ←	<b>0.65</b>	0.70 ←	<b>1.65</b>	1.00 ←	<b>1.28</b>
	<b>Labials</b>	0.90 ←	<b>1.25</b>	0.19 ←	<b>0.77</b>	0.97 ←	<b>1.37</b>
	<b>Coronals</b>	1.17 →	0.94	1.25 →?	1.24	1.06 →	0.68

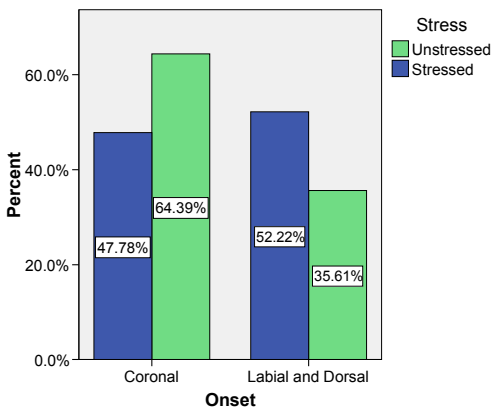
- (27) There appears to be a connection between the quality of the consonant and stress. A chi-square between onset place (collapsing labials and dorsals) and stress:
- (a) In CMU:  $\chi^2(1, N = 84,346) = 2843.4, p < .001$ .
  - (b) In CELEX:  $\chi^2(1, N = 25,888) = 702.12, p < .001$ .
- (28) A chi-square between coda place (collapsing labials and dorsals) and stress:
- (a) In CMU:  $\chi^2(1, N = 84,346) = 478.02, p < .001$ .
  - (b) In CELEX:  $\chi^2(1, N = 25,888) = 47.74, p < .001$ .

Table 11: Onset \* Stress crosstabulation (CELEX).

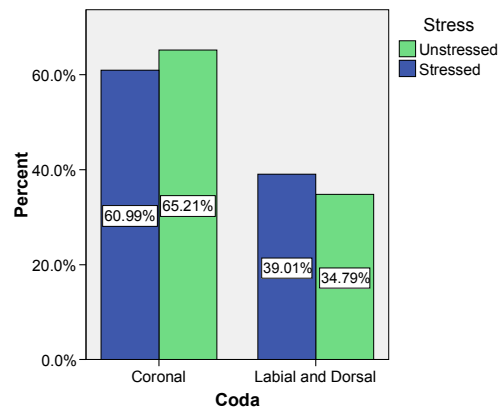
		Stress		Total
		Unstressed	Stressed	Unstressed
Onset	Unmarked	10008	4943	14951
	Marked	5535	5402	10937
Total		15543	10345	25888

- (29) The strength of the relationship, measured by Phi and Cramer's V, was higher for onset-stress than for coda-stress in both dictionaries (0.184 and 0.165 vs. 0.075 and 0.043).
- (30) Unstressed syllables prefer coronal onsets, stressed syllables prefer labial or dorsal onsets.
- (31) Unstressed syllables prefer coronal codas; stressed syllables prefer labial or dorsal codas.

Graph 3: Onset place and stress in CELEX



Graph 4: Coda place and stress in CELEX



- (32) Constraints:

\*x/p\_  
 \*X/t\_  
 \*x/\_p  
 \*X/\_t

Avoid unstressed syllables with labial/dorsal onsets.  
 Avoid stressed syllables with coronal onsets.  
 Avoid unstressed syllables with labial/dorsal codas.  
 Avoid stressed syllables with coronal codas.

- (33) Onset-stress constraints are ranked higher than coda-stress constraints.
- (34) These constraints directly connect consonant place and stress. This runs counter to the proposal that place and stress do not directly interact (see e.g. de Lacy 2001, Blumenfeld 2006). An alternative account based on faithfulness constraints is imaginable, but left for future work.

## 6. Interim summary

- (35) We propose a harmonic alignment hypothesis: syllable prominence scale in English is aligned with consonant place and vowel sonority scales.

X > x	stressed syllables are more prominent than unstressed
A > I	low vowels are more sonorous than high and reduced vowels
p, k > t	labial and dorsal consonants are more marked/prominent than coronal

X/A > x/A
x/I > X/I

X/p > x/p
x/t > X/t

- (36) Each syllable is coded for four factors:
- Type of onset: marked (dorsal and labial) and unmarked (coronal).
  - Type of coda: marked (dorsal and labial) and unmarked (coronal).
  - Type of vowel: low (all non-high vowels) and high (all high vowels and schwa).
  - Type of syllable: stressed and unstressed.

- (37) Summary of constraints:

OCP	Avoid identical place features in the onset and coda.
OCP-marked	Avoid identical marked place features in the onset and coda.
*x/p_	Avoid unstressed syllables with labial/dorsal onsets.
*X/t_	Avoid stressed syllables with coronal onsets.
*x/_p	Avoid unstressed syllables with labial/dorsal codas.
*X/_t	Avoid stressed syllables with coronal codas.
*X/I	Avoid stressed syllables with high vowels.
*x/A	Avoid unstressed syllables with low vowels.

- (38) Rankings:

- (a) Vowel-stress constraints are ranked above consonant-stress (onset-stress, coda-stress) constraints.

\*x/A, \*X/I >> \*x/p\_, \*x\_p, \*X/t\_, \*X/\_t

- (b) Onset related constraints are ranked above coda-related constraints.

\*x/p\_, \*X/t\_ >> \*x/\_p, \*X/\_t

(39) Table 12: Matching pattern of constraint violations with O/E values in CMU (ordered by ascending O/E values).

Onset	Coda	Vowel	Stress	Example	Observed frequency	Expected frequency	O/E value	Constraint Violations
labial	labial	high	unstressed	[bi:f]	37	838.062	0.044	4
dorsal	dorsal	low	unstressed	[hæk]	24	333.8999	0.072	5
labial	labial	low	unstressed	[bob]	40	319.0057	0.125	5
coronal	coronal	low	unstressed	[tol]	1009	6767.922	0.149	2
labial	labial	high	stressed	[bi:p]	107	442.5329	0.242	3
dorsal	coronal	low	unstressed	[kon]	461	1781.01	0.259	2
dorsal	dorsal	high	stressed	[kIk]	120	463.1945	0.259	3
coronal	dorsal	low	unstressed	[leg]	427	1268.836	0.337	2
coronal	labial	low	unstressed	[top]	223	654.3556	0.341	2
labial	coronal	low	unstressed	[bel]	1148	3299.438	0.348	2
coronal	coronal	high	stressed	[lut]	3294	9388.634	0.351	4
coronal	dorsal	high	stressed	[dIg]	644	1760.161	0.366	2
labial	dorsal	low	unstressed	[peg]	230	618.5717	0.372	3
dorsal	dorsal	high	unstressed	[kIŋ]	404	877.1905	0.461	4
dorsal	coronal	high	stressed	[kIt]	1168	2470.662	0.473	2
labial	dorsal	high	stressed	[bIg]	409	858.0985	0.477	1
dorsal	labial	high	stressed	[gIv]	145	238.8756	0.607	1
coronal	labial	high	stressed	[zIp]	588	907.7389	0.648	2
labial	coronal	high	stressed	[bul]	3226	4577.065	0.705	2
coronal	labial	high	unstressed	[tIv]	1754	2343.552	0.748	1
labial	dorsal	high	unstressed	[wIg]	1399	1625.054	0.861	2
dorsal	labial	low	unstressed	[kom]	149	172.1967	0.865	3
dorsal	labial	high	unstressed	[kəm]	430	452.379	0.951	2
dorsal	coronal	high	unstressed	[gud]	5257	4678.901	1.124	1
labial	coronal	high	unstressed	[mIs]	10448	8667.974	1.205	1
coronal	coronal	high	unstressed	[tən]	26006	17780.05	1.463	1
coronal	coronal	low	stressed	[lot]	5283	3573.755	1.478	3
coronal	dorsal	high	unstressed	[dIŋ]	5643	3333.366	1.693	1
dorsal	dorsal	low	stressed	[ho:k]	356	176.3136	2.019	2
labial	labial	low	stressed	[mæm]	367	168.4488	2.179	2
coronal	dorsal	low	stressed	[dog]	1649	670	2.461	1
labial	dorsal	low	stressed	[bæk]	825	326.6326	2.526	0
labial	coronal	low	stressed	[wel]	5360	1742.246	3.076	1
dorsal	coronal	low	stressed	[hot]	3301	940.45	3.510	1
coronal	labial	low	stressed	[dæm]	1576	345.528	4.561	1
dorsal	labial	low	stressed	[hæm]	839	90.92729	9.227	0



(40) Table 13: Matching pattern of constraint violation with O/E values in CELEX (ordered by ascending O/E values).

Onset	Coda	Vowel	Stress	Example	Observed frequency	Expected frequency	O/E value	Constraint Violations
labial	labial	high	unstressed	[bi:f]	27	404.5	0.067	4
dorsal	dorsal	high	stressed	[kIk]	40	203.0	0.197	3
coronal	coronal	low	unstressed	[tol]	472	2178.7	0.217	2
dorsal	dorsal	low	unstressed	[hæk]	49	188.6	0.260	5
dorsal	coronal	high	stressed	[kIt]	187	616.7	0.303	2
coronal	coronal	high	stressed	[lut]	777	2344.9	0.331	4
coronal	labial	low	unstressed	[top]	177	533.9	0.331	2
coronal	dorsal	low	unstressed	[leg]	251	717.3	0.350	2
labial	labial	high	stressed	[bi:p]	95	269.3	0.353	3
dorsal	labial	high	stressed	[gIv]	55	151.1	0.364	1
labial	labial	low	unstressed	[bob]	94	250.2	0.376	5
coronal	dorsal	high	stressed	[dIg]	322	772.0	0.417	2
labial	dorsal	high	stressed	[bIg]	202	361.7	0.558	1
labial	coronal	low	unstressed	[bel]	639	1020.8	0.626	2
dorsal	labial	low	unstressed	[kom]	89	140.4	0.634	3
dorsal	coronal	low	unstressed	[kon]	374	573.0	0.653	2
coronal	labial	high	stressed	[zIp]	385	574.7	0.670	2
labial	coronal	high	stressed	[bul]	802	1098.7	0.730	2
dorsal	labial	high	unstressed	[kəm]	169	227.1	0.744	2
dorsal	dorsal	high	unstressed	[kIŋ]	236	305.1	0.774	4
labial	dorsal	low	unstressed	[peg]	267	336.1	0.794	3
labial	dorsal	high	unstressed	[wIg]	520	543.5	0.957	2
labial	coronal	high	unstressed	[mIs]	1946	1650.7	1.179	1
dorsal	coronal	high	unstressed	[gud]	1125	926.6	1.214	1
coronal	coronal	low	stressed	[lot]	1815	1450.1	1.252	3
dorsal	dorsal	low	stressed	[ho:k]	172	125.6	1.370	2
labial	labial	low	stressed	[mæm]	241	166.5	1.447	2
coronal	coronal	high	unstressed	[tən]	5579	3523.1	1.584	1
coronal	dorsal	high	unstressed	[dIŋ]	1954	1159.9	1.685	1
coronal	dorsal	low	stressed	[dog]	869	477.4	1.820	1
coronal	labial	high	unstressed	[tIv]	1575	863.4	1.824	1
coronal	labial	low	stressed	[dæm]	775	355.4	2.181	1
labial	dorsal	low	stressed	[bæk]	532	223.7	2.378	0
labial	coronal	low	stressed	[wel]	1640	679.4	2.414	1
dorsal	coronal	low	stressed	[hot]	1088	381.4	2.853	1
dorsal	labial	low	stressed	[hæm]	348	93.5	3.723	0

Observations:

- (a) Shaded vowel/stress cells show the candidates that incur \*X/I or \*x/A violations.
- (b) Shaded onset/coda cells show the candidates that incur OCP violations.

In both cases the penalized candidates are clustered in the top of the table. They are also the most underrepresented candidates.

- (41) A priori rankings mean that not only the absolute number of violations matters, but also the kind of constraint violated.

### 7. Testing the grammar

- (42) Hypothesis: The relative well-formedness of a phonotactic combination depends on its GRAMMATICAL COMPLEXITY in the following sense: the more ranking information a phonotactic combination requires in order to surface faithfully, the less well-formed it is.

- (43) Example: Compare two CVC syllables:

- (a) [<sup>h</sup>bIɡ] labial onset, high stressed vowel, dorsal coda
- (b) [<sup>h</sup>kIk] dorsal onset, high stressed vowel, dorsal coda

- (44) In order to achieve a faithful mapping, the following rankings are required:

- (a) <[<sup>h</sup>bIɡ], [<sup>h</sup>bIɡ]> FAITH >> \*X/I O/E = 0.477 / 0.558
- (b) <[<sup>h</sup>kIk], [<sup>h</sup>kIk]> FAITH >> \*X/I, FAITH >> OCP O/E = 0.259 / 0.197

/ <sup>h</sup> bIɡ/	*OCP	Faith	*X/I
☞ [ <sup>h</sup> bIɡ]			*
OTHER		*	

/ <sup>h</sup> kIk/	Faith	*OCP	*X/I
☞ [ <sup>h</sup> kIk]		*	*
OTHER	*		

- (45) The second mapping is more complex than the first in the sense that it ENTAILS the first:

<[<sup>h</sup>bIɡ], [<sup>h</sup>bIɡ]> → <[<sup>h</sup>kIk], [<sup>h</sup>kIk]>

- (46) The Complexity Hypothesis: The probability of an <input, output> mapping is inversely correlated with its grammatical complexity.

- (47) A T(YPOLOGICAL) ORDER is the set of all ranking entailments derived by the grammar.

- (48) The T-order (CELEX data) for the grammar outlined above:



(53) Asymmetry 2:

- Stressed syllables prefer labial and dorsal (marked) onsets and codas.
- Unstressed syllables prefer coronal (unmarked) onsets and codas.
- In both cases this preference is stronger for onsets than for codas.

## References

- Anttila, Arto and Curtis Andrus. 2006. T-Orders. Stanford University, ms. Available online at <http://www.stanford.edu/~anttila/research/software.html>.
- Baayen, R. Harald, Richard Piepenbrock, and Leon Gulikers. 1995. The CELEX Lexical Database (Release 2). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].
- Berkley, Deborah. 1994. The OCP and gradient data. *Studies in the Linguistic Sciences* 24.59–72.
- Blumenfeld, Lev. 2006. Constraints on phonological interactions. Stanford University: PhD Dissertation. Available on ROA.
- Coetzee, Andries. 2004. What it means to be a loser: non-optimal candidates in Optimality Theory. University of Massachusetts, Amherst: Ph.D. Dissertation.
- Frisch, Stefan A., Janet B. Pierrehumbert, and Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22(1).179-228.
- Goldsmith, John. 1976. *Autosegmental Phonology*. Cambridge, MA: MIT dissertation.
- Hayes, Bruce, Bruce Tesar, and Kie Zuraw. 2003. OTSoft 2.1, software package. Available online at <http://www.linguistics.ucla.edu/people/hayes/otsoft/>.
- Kenstowicz, Michael. 1994. Sonority-driven stress. MIT: Ms. ROA #33.
- de Lacy, Paul. 2001. Markedness in prominent positions. In Ora Matushansky, Albert Costa, Javier Martin-Gonzalez, Lance Nathan, and Adam Szczegielniak (eds.) *HUMIT 2000*, MIT Working Papers in Linguistics 40. Cambridge, MA: MITWPL, pp.53-66. [ROA #432](#).
- Leben, William. 1973. *Suprasegmental phonology*. Cambridge, MA: MIT dissertation.
- McCarthy, John J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17.207–263.

Odden, David. 1986. On the role of the Obligatory Contour Principle in phonological theory. *Language* 62.353–383.

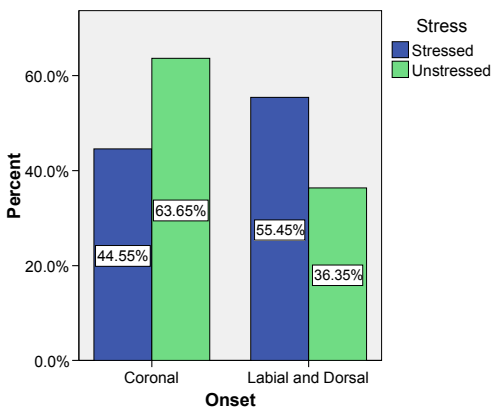
Prince, Alan & Smolensky, Paul. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell. [Revision of 1993 technical report, RUCCS. Available on Rutgers Optimality Archive, ROA-537].

Weide, Robert. 1998. *The CMU pronunciation dictionary (Release 0.6)*. Carnegie Mellon University. Available online at <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>.

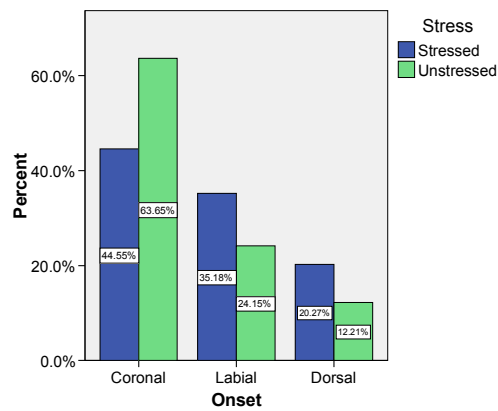
Yip, Moira. 1988. The Obligatory Contour Principle and phonological rules: a loss of identity. *Linguistic Inquiry* 19.65–100.

**Appendix A: More graphs for onset-stress and coda-stress relationship in CMU and CELEX.**

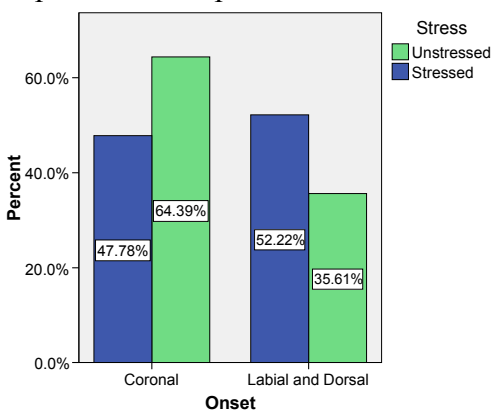
Graph 1.1: Onset place and stress in CMU



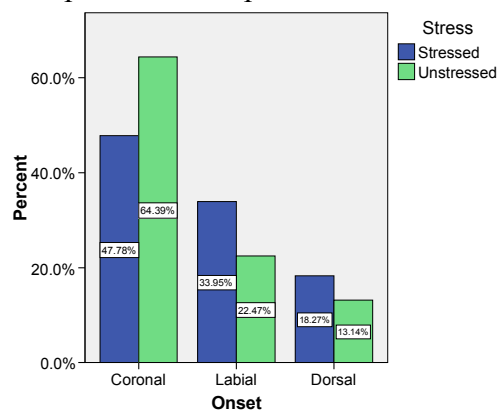
Graph 1.2: Onset place and stress in CMU



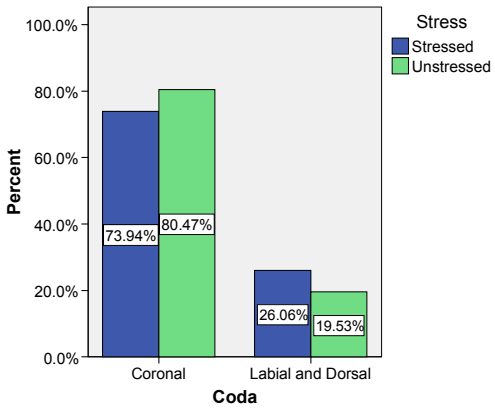
Graph 2.1: Onset place and stress in CELEX



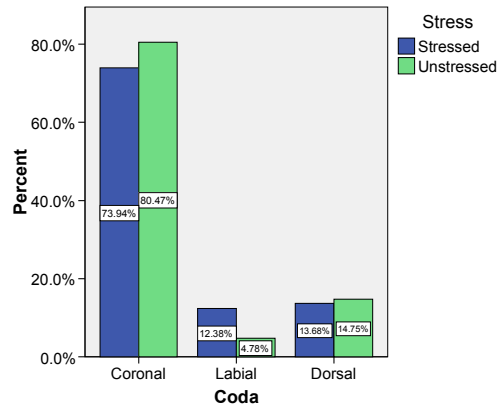
Graph 2.2: Onset place and stress in CELEX



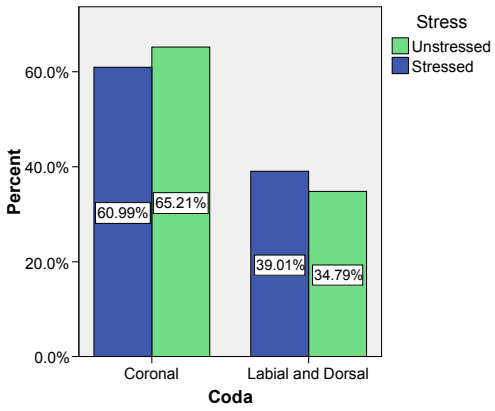
Graph 3.1: Coda place and stress in CMU



Graph 3.2: Coda place and stress in CMU



Graph 2.1: Coda place and stress in CELEX



Graph 2.2: Coda place and stress in CELEX

