Gradient Lexical Reflexes of the Syllable Contact Law*

Kevin B. McGowan
University of Michigan

1 Introduction

The Syllable Contact Law (SCL) predicts a gradient preference for syllable contact pairs in a language. However, this law is most often implemented with strictly categorical constraints. In this paper I argue that the gradience underlying the SCL operates synchronically in the grammar of a language. I examine CELEX corpus data for evidence of gradient lexical reflexes of the SCL in British English using pointwise mutual information (PMI). The evidence appears to support a gradient hypothesis and to be inconsistent with a strictly categorical model.

2 The Syllable Contact Law

Vennemann (1988) and Murray & Vennemann (1983) propose the SCL which attempts to explain syllabification patterns and sound change at syllable boundaries in terms of a single, graded preference "law". The proposed law can be paraphrased:

> A syllable contact pair $\alpha.\beta$ is more preferred the greater the increase in consonantal strength from a coda segment α to an onset segment β.

Consonantal strength, according to Vennemann, is the degree to which the airflow of a speech sound deviates from otherwise unimpeded airflow. The SCL defined this way predicts that syllable boundaries are preferred if the consonantal strength at the end of the coda is lower than the consonantal strength of the following onset. In other words, airflow during the coda of a syllable should be as unimpeded as possible while airflow at the onset of the following syllable should be as impeded as possible. Vennemann (1988) provides a ranking (see below) of speech sounds in terms of their consonantal strength.

From either a diachronic or synchronic perspective (Vennemann addresses these separately), the SCL makes an explicitly gradient prediction: syllable contact pairs across which consonantal strength minimally increases[1] will be the first to be affected by sound change. Furthermore, a transitive relationship holds across grammaticality judgments: if contact pair A is preferred to B and B is preferred to C then A will also be preferred to C (and to a greater extent than B).

*I owe thanks to the audience at MCWOP 14 for questions and general discussion of an earlier version of this work. I owe particular thanks to Steven Abney, Patrice S. Beddor, Andries Coetzee and San Duanmu for their generous guidance and suggestions. Finally, thanks to Benjamin Munson for sharing Pierrehumbert's (2001) list of monomorphemic CELEX words with me.

[1]Note that this minimal increase may actually be a decrease.

2.1 Sonority and the SCL

Recent discussions of the SCL (e.g. Davis 1998, Parker 2002, Gouskova 2004, etc.) have remained largely faithful to Vennemann's wording of the law but replace the evaluation metric; restating the law in terms of sonority. Sonority and consonantal strength are portrayed as merely inversions of one another, but this is not exactly true. Gouskova, for example, adopts Jespersen's (1904) sonority hierarchy which, unlike Vennemann's scale, interleaves the voiced and voiceless fricatives with the voiced and voiceless stops. This paper will follow Gouskova in using Jespersen's sonority hierarchy so it may be useful to compare it to the scale of consonantal strength Vennemann provides.

consonantal strength scale rhotics $\prec$ laterals $\prec$ nasals $\prec$ voiced fricatives $\prec$ voiceless fricatives $\prec$ **voiced stops** $\prec$ **voiceless stops**

sonority hierarchy glides $\succ$ rhotics $\succ$ laterals $\succ$ nasals $\succ$ voiced fricatives $\succ$ **voiced stops** $\succ$ voiceless fricatives $\succ$ **voiceless stops**

Sonority has long been a source of debate among phonologists and phoneticians (cf. Ohala 1990). Parker 2002 quantifies the sonority along acoustic, articulatory, and aerodynamic dimensions. He finds a .97 correlation between traditional sonority hierarchies (like Jespersen's) and measurements, listed in ranked order of contribution to the linear model, of:

1. intensity (positively correlated)
2. peak intraoral air pressure (negatively correlated)
3. F1 frequency
4. peak airflow and
5. total segment duration.

Vennemann's definition of consonantal strength as impedance of airflow can be loosely modeled using both the peak intraoral air pressure and peak airflow features in Parker's study[2], these two scales can arguably be treated as rough inversions of one another[3]. When evaluating the predictions made by Vennemann and others, though, it seems useful to bear the scale permutations and metric mismatches in mind. With these caveats, the present paper will follow the recent models and use a version of the SCL restated in terms of this sonority hierarchy:

> A syllable contact pair $\alpha.\beta$ is more preferred the greater the drop in sonority from a coda segment α to an onset segment β.

Ideal syllable contact pairs will tend to maximize the first differences from coda to onset in terms of sonority. Figure 2.1 shows a schematic of such an idealization. This maximal differentiation along a physiological dimension makes intuitive sense as a cue to the perception of syllabification and has received much attention

[2]Actually, none of these features is independent; though it should be noted that airflow alone accounted for very little of the variation in Parker's model.

[3]Davis (1998) goes further and suggests that sonority is the more phonetically motivated of the two metrics.

in the phonetics and phonology literature (cf. Krakow 1999). Viewed from this perspective, one could interpret the words 'more preferred' in the SCL formulation as 'more perceptible' or, perhaps, 'more perceptually salient'.

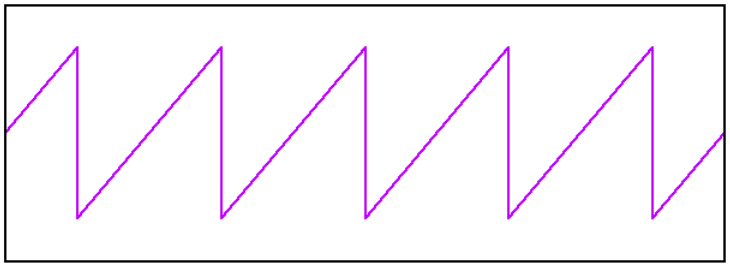

Figure 1: idealization of preferred sonority change across syllable boundaries.

2.2 A Gradient Process or Categorical Rule?

Davis (1998) and Gouskova (2004) each argue, albeit with strikingly different outcomes, against a gradient synchronic interpretation of the SCL. Both note that a single, strictly categorical SCL constraint is sufficient to account for data from Kazakh in which maximization of sonority difference is not required and any drop in sonority is sufficient to satisfy the constraint. Evidence of languages in which maximization is required (e.g. Kirghiz), however, and the need to account for cross-linguistic differences (e.g. degree of tolerated sonority rise in Icelandic and Faroese) lead Gouskova to proffer a set of relational hierarchical constraints.

Using Jespersen's sonority hierarchy, Gouskova extrapolates the coda and onset sonority rankings in Figure 2. These motivate the predicted syllable contact hierarchy in Table 2.2. The scale across the bottom of this table represents sonority change; a sonority distance of -7 represents the best possible syllable-contact pairing of consonants in a language of the world while a distance of +7 represents the worst possible pairing. The presences of a +4 pairing in a language implies the existence in that language of +3 down to -7 but not of +5 or higher.

coda $w \succ r \succ l \succ n \succ z \succ d \succ s \succ t$

onset $t \succ s \succ d \succ z \succ n \succ l \succ r \succ w$

Figure 2: onset and coda hierarchies in Gouskova (2004)

Central to the argument here is that while these relational, hierarchical constraints are sufficient to capture cross-linguistic differences in sonority slope, it is not possible to derive gradient grammaticality judgments from them. Coetzee & Pater (2008) identify an identical problem with categorical constraints and, for example, the intermediate attestedness of Arabic fricative/stop pairs. Neither a candidate form with a sonority distance of +4 nor a candidate with a sonority distance of -2 will violate a high ranking constraint against sonority distances of +5 or greater. No permutation of these simple, atomic constraints can account for changes of type -2 being more prevalent in the language than changes of type +4. A language with a cut-off point of +5 should, all other things being equal, provide categorical grammaticality judgments below that level and a uniform distribution of lexical items by syllable contact distance.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
w.t	w.s	w.d	w.z	w.n	w.l	w.r	w.w	r.w	l.w	n.w	z.w	d.w	s.w	t.w
	r.t	r.s	r.d	r.z	r.n	r.l	r.r	l.r	n.r	z.r	d.r	s.r	t.r	
		l.t	l.s	l.d	l.z	l.n	l.l	n.l	z.l	d.l	s.l	t.l		
			n.t	n.s	n.d	n.z	n.n	z.n	d.n	s.n	t.n			
				z.t	z.s	z.d	z.z	d.z	s.z	t.z				
					d.t	d.s	d.d	s.d	t.d					
						s.t	s.s	t.s						
							t.t						= observed	
-7	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6	+7

Table 1: Gouskova (2004) Sonority Contact Scale (with CELEX observed pairs)

IPA	Sonority	IPA	Sonority
p	8	s	7
b	6	z	5
t	8	ʃ	7
d	6	ʒ	5
k	8	j	1
g	6	h	7
ŋ	4	w	1
m	4	ʧ	8
n	4	ʤ	6
l	3	ŋ̍	4
ɹ	2	m̩	4
f	7	n̩	4
v	5	l̩	3
θ	7	ɹ̩	2
ð	5		

Table 2: Sonority values assigned to each phoneme to calculate sonority distance.

The grey boxes in Table 2.2 indicate pair types actually observed in the CELEX British English corpus (Baayen *et al.* 1995). The fact that, for example, no r.C pairs appear in the database is not a problem for a categorical implementation of the SCL. The non-occurrence of r.C in the CELEX corpus is a consequence of the particular dialect of English documented therein. An account like Gouskova's would simply include a markedness constraint against rhotics in coda position (e.g. *CODA/R). With the exception of cases like these, Gouskova's categorical model predicts a uniform distribution of syllable contact pairs or, at the very least, does not predict a distribution reflecting the preferences of the SCL.

3 The SCL in CELEX data

The question to be answered, then, is whether the gradient transitive relationship among lexical items predicted by the SCL holds for some sample of language data or, if not, whether the distribution of those data more closely reflects the outcome of simple, categorical grammaticality judgments. However, we cannot simply sum the syllable contact tokens at each sonority distance level and compare the counts directly. Among other problems with this approach, there are eight times as many contact pair types at sonority distance 0 than at the extremes of -7 or +7. A simple summing of tokens from this system, assuming pairs are generated uniformly at random, will approximate a normal distribution for precisely the same reason that throwing a pair of fair dice multiple times will.

3.1 Pointwise Mutual Information

What crucially differentiates a pattern consistent with a gradient interpretation from a categorical interpretation is not simple frequency of a syllable contact type in a lexicon of the language but the extent to which the occurrence of syllable contact pairings differs from the expected pairings of those phonemes if they were independently distributed. Under a gradient interpretation, the contact types Vennemann describes as 'more preferred' should be more highly associated with one another than those contact types which are less preferred. A more uniform distribution below the grammatical cut-off point in a language, on the other hand, would be more consistent with a categorical interpretation.

A simple χ^2 test captures this relationship of observed to expected values and could allow us to judge the probability of individual pairings given the underlying distribution. A more fitting metric for this experiment, though, is Pointwise Mutual Information or PMI.

PMI (Church & Hanks 1989) is a metric from computational and corpus linguistics that has often been used to identify collocations (strongly associated pairs of words) in corpora and to link words with their probable translations in parallel bilingual texts. Both χ^2 and PMI tell us something about the independence of two variables, but PMI also, importantly, tells us something about the *dependence* of those two variables. PMI, which is expressed in terms of bits of information, indicates the strength of association between two phonemes. Positive PMI values indicate that two phonemes tend to occur together, values close to zero suggest independence, and negative values indicate that one or both of the phonemes tend to occur where the other does not.

In the following equation, $p(\alpha)$ and $p(\beta)$, represent the maximum likelihood estimate for each segment in a candidate contact pair. The PMI of two segments, then, is the $\log_2$ of the ratio of the joint probability of those two segments to their probability assuming independence.

$$(1) \quad PMI(X=\alpha, Y=\beta) = \log_2 \frac{P(\alpha,\beta)}{P(\alpha)P(\beta)}$$

PMI is conceptually similar to an increasingly common metric in the phonology literature, O/E ratios (used by, e.g., Pierrehumbert 1992, Frisch 1996, Frisch *et al.* 2004 and Coetzee & Pater 2008). Whereas PMI compares the observed frequency of two items to their frequency assuming independence, O/E ratios compare observed frequency to an expected, random distribution. PMI asks the question "are these items independent/dependent?" while O/E ratios ask the question "how does the frequency of this pair differ from expected frequency if pairs are uniformly distributed?".

3.2 CELEX

The British English portion of the CELEX database was chosen because it includes a broad phonetic transcription, syllabification information, and morphemic status for each entry. Polymorphemic words were excluded because the SCL is a prediction about syllable, rather than morpheme, boundaries. However, the CELEX database has only 7,401 words explicitly labeled monomorphemic (of which only 815 are polysyllabic). If one is willing to accept a word list that is simultaneously too conservative and too liberal in its exclusion and inclusion of genuinely monomorphemic words, CELEX has 10,738 words labelled either monomorphemic or 'obscure'. Fortunately, Pierrehumbert (2001) manually identified 11,383 monomorphemic words in English CELEX though only 2,002 of these words are also polysyllabic. A perl script calculated PMI for each observed segment pair in the monomorphemic, polysyllabic CELEX words following the approach described in Manning & Schütze (1999).

Each phoneme was assigned an inherent sonority value as shown in Table 2 and the sonority distance for each contact pair was calculated by subtracting the sonority value of the onset consonant from the sonority value of the coda consonant. A word like *new.ton*[4] would therefore have a single contact pair with a sonority distance of -7 (the best possible contact type according to the SCL) while the contact pair in *nit.wit* would have a distance of +7 (the worst possible contact type).

3.3 Results

PMI	Sonority Change	$\alpha.\beta$	PMI	Sonority Change	$\alpha.\beta$
7.15	-2	ŋ-g	3.38	-4	n-t
4.93	-4	ŋ-k	3.38	1	z-m
4.39	-2	m-b	3.18	-3	ŋ-h
4.27	0	f-θ	2.91	-1	n-v
4.23	-2	n-d	2.91	-3	n-s
4.16	-4	m-p	2.71	2	g-n
3.84	1	g-z	2.62	6	θ-w
3.73	1	k-ʃ	2.60	-3	n-ʃ
3.69	-2	n-ʤ	2.51	1	ð-m
3.47	-3	n-θ	2.50	2	g-m

Table 3: 20 most strongly associated contact pairs

The 20 contact pairs most highly ranked by PMI are shown in Table 3.3. One readily apparent outcome is that five of the top six associations, at least, involve a

[4]If transcribed with the voiced labial-velar approximant in coda position of the first syllable.

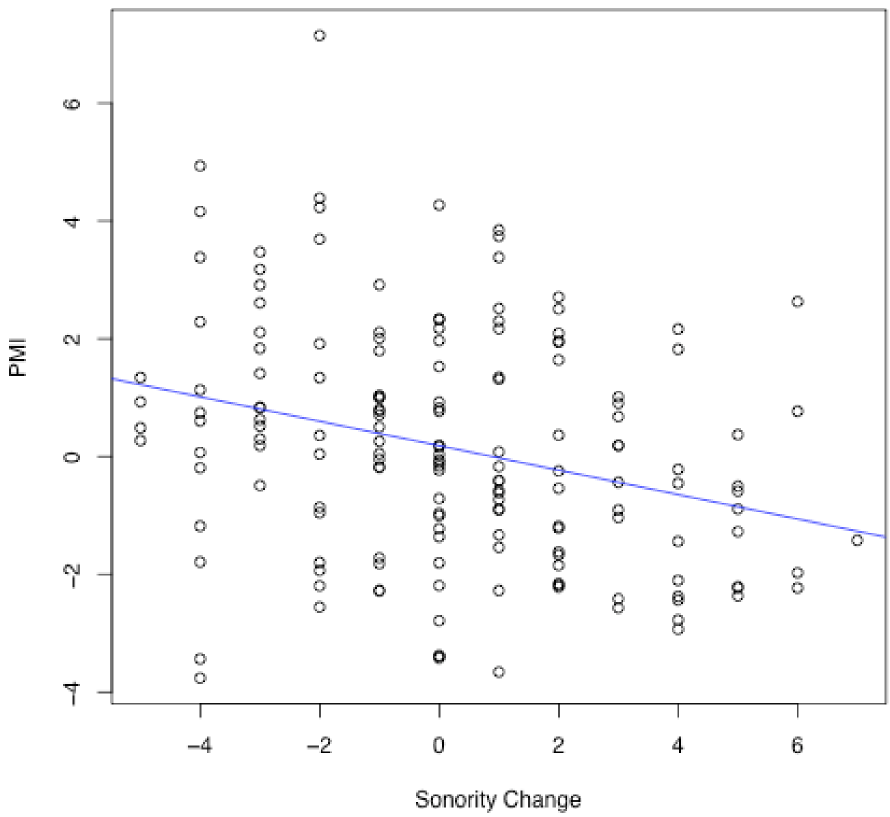

Figure 3: PMI by sonority change

homorganic nasal-plosive pairing. Consistencies of this sort were, after all, Venneman's motivation for the SCL: explaining patterns of diachronic change at syllable boundaries. One might, for example, note the strong association of these pairs and look for evidence of a consistent pattern of nasal-plosive place assimilation throughout the lexicon or look for evidence of English listeners' ability to perceive a difference in place between nasal-plosive syllable contact pairs.

However, sonority distance does not seem to be the organizing factor in these results. One might expect if the SCL were the primary organizing principle for syllable contact pairs and if PMI accurately captured the strength of association across contact pairs that a ranking in terms of PMI would also be a ranking in terms of sonority change. However, the SCL is not really a law about any particular pair of phonemes which may be biased any number of other constraints: coarticulation, dissimilation, lexical neighborhood density etc. A fairer test of the SCL is to evaluate how well it predicts patterns of syllable contact across the entire lexicon.

An interesting picture emerges when we plot the full set of PMI data by sonority distance as in Figures 3.3 & 3.3. Not only is there a clear slope in the data showing the expected transitive relationship (the steeper the sonority drop, the better the pair), but note the complete absence of data points in the upper right-hand corner of

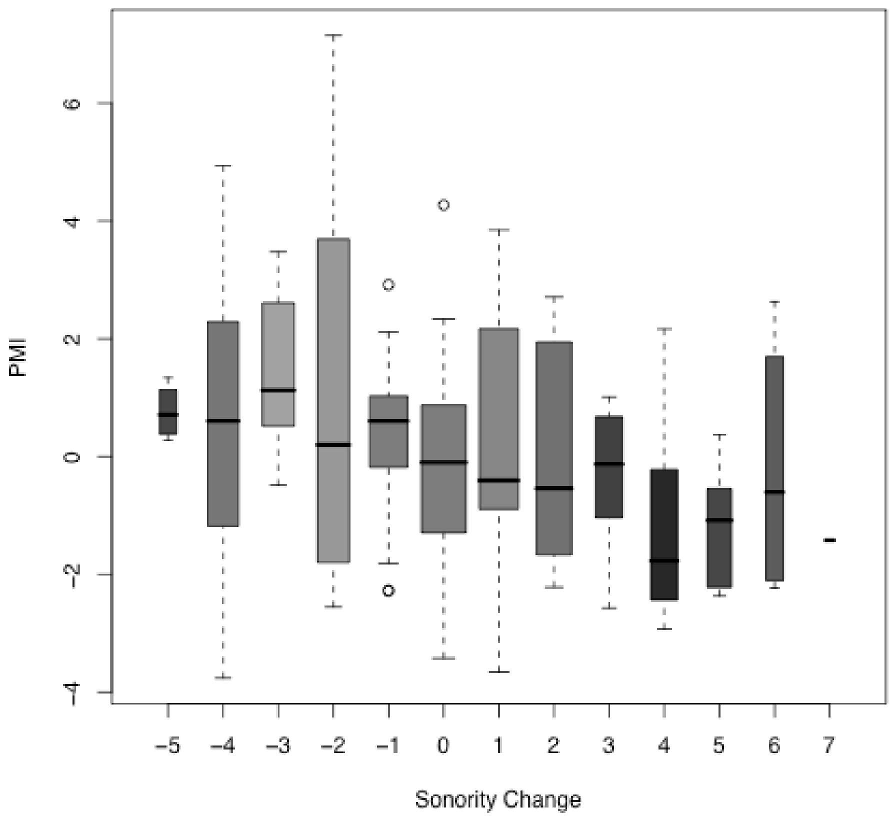

Figure 4: PMI boxplot by sonority change

the plot. Contact pairs with a steep sonority drop attract one another while contact pairs with a sonority rise repel one another. This pattern is consistent with a gradient interpretation of the SCL.

Furthermore, we can see in Figure 3.3 that the median PMI is positive only for drops in sonority. The median PMI for sonority distances of 0 or greater are all negative. Only contact pairs for which sonority falls are positively associated with one another; contact pairs with a positive sonority slope occur less frequently than they would independently. The linear model plotted in Figure 3.3 shows a significant interaction of PMI and sonority distance (p = 0.0002). The model falls far short of explaining all of the variation in the data (r^2 = .077), but syllable contact pairs chosen uniformly at random would not show this same slope.

4 Alternative Interpretation: Synchrony vs. Diachrony

There is, at least, one alternative interpretation of the results of this corpus analysis that the present work can neither confirm nor refute. The CELEX data may still be compatible with categorical/non-gradient grammars if we argue that gradient patterns in the lexicon are merely the residue of the graded application of diachronic rules.

Placing diachronic change outside the synchronic grammar of any individual speaker seems to suggest, however, that diachronic change is an emergent property arising either from the microdynamics of a speech community or from the nature of the mental lexicon. Within such a model, individual speakers' grammatical judgments can be simple and categorical with gradient lexical reflexes emerging from, for example, patterns of perceptual saliency (cf. Ohala 1981), gestural variability (cf. Browman & Goldstein 1991) or both (cf. Lindblom 1990).

This is a plausible explanation but is well beyond the scope of the present work. It may be possible to investigate this hypothesis with a psycholinguistic experiment testing grammaticality judgments, but even this approach seems difficult to disentangle from the subjects' awareness of statistical patterns in their native language(s) like those visible in CELEX.

5 Conclusion

The goal of this paper has been to compare the predictions made by categorical and gradient implementations of the syllable contact law with data in a lexicon of a language. The data in the British English CELEX corpus does indeed show, as Gouskova's model would predict, that the existence of the most negatively marked contact pair type implies (with the caveats noted above) the existence of all less marked types. This finding is perfectly consistent with a categorical model in which each language has a language-specific cut-off above which syllable contact pairs are ungrammatical.

However, the CELEX data also show that syllable contact pairs with a sonority decrease are positively associated while syllable contact pairs with a sonority increase are negatively associated (and syllable contact pairs with a sonority distance of 0 are nearly independent). This finding is consistent with a gradient implementation of the syllable contact law but is inconsistent with a categorical implementation. It is likely that an application of Harmonic Grammar such as that worked-out for Muna and Arabic by Coetzee & Pater (2008) could better accommodate the frequency patterns of English syllable contact pairs.

References

Baayen, R., R. Piepenbrock, & L. Gulikers. 1995. The celex lexical database. CD-ROM. Philadelphia: Linguistic Data Consortium.

Browman, C., & L. Goldstein. 1991. Gestural structures: distinctiveness, phonological processes, and historical change. In *Modularity and the motor theory of speech perception*, ed. by I.G. Mattingly & M. Studdert-Kennedy, 313–338. Laurence Erlbaum Associates.

Church, Kenneth W., & Patrick Hanks. 1989. Word association norms, mutual information, and lexicography. In *Proceedings of the 27th. Annual Meeting of the Association for Computational Linguistics*, 76–83, Vancouver, B.C. Association for Computational Linguistics.

Coetzee, Andries W., & Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in muna and arabic. *Natural Language and Linguistic Theory* 26.289–337.

Davis, Stuart. 1998. Syllable contact in optimality theory. *Korean Journal of Linguistics* 23.181–211.

Frisch, Stefan, 1996. *Similarity and Frequency in Phonology*. Northwestern University dissertation.

Frisch, Stefan A., Janet B. Pierrehumbert, & Michael B. Broe. 2004. Similarity avoidance and the ocp. *Natural Language and Linguistic Theory* 22.179–228.

Gouskova, Maria. 2004. Relational hierarchies in optimality theory: the case of syllable contact. *Phonology* 201–250.

Jespersen, Otto. 1904. *Lehrbuch der Phonetik*. Leipzig and Berlin: Teubner.

Krakow, Rena A. 1999. Physiological organization of syllables: a review. *Journal of Phonetics* 27.23–54.

Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H & H theory. In *Speech production and speech modeling*, ed. by Marchal Hardcastle, 403–439. Dordrecht: Kluwer.

Manning, Christopher D., & Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Murray, Robert W., & Theo Vennemann. 1983. Sound change and syllable structure in germanic phonology. *Language* 59.514–528.

Ohala, J.J. 1981. The listener as a source of sound change. In *Papers from the Parasession on Language and Behavior*, ed. by C.S. Masek, R.A. Hendrik, & M.F. Miller, 178–203. Chicago Linguistic Society.

Ohala, John J. 1990. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In *CLS 26: Volume 2: the Parasession on the syllable in phonetics and phonology.*, ed. by Michael Ziolkowski, Manual Noske, & Karen Deaton, 319–338. Chicago: Chicago Linguistic Society.

Parker, Stephen, 2002. *Quantifying the Sonority Hierarchy*. University of Massachusetts Amherst dissertation.

Pierrehumbert, Janet. 1992. Dissimilarity in the arabic verbal roots. In *Proceedings of the 23rd Meeting of the Northeastern Linguistic Society*, 367–381. Graduate Student Association, U. Mass. Amherst.

——. 2001. Why phonological constraints are so coarse grained. In *SWAP (Spoken Word Access Processes) Special Issue: Language and Cognitive Processes 16:5*, ed. by J. M. McQueen & A. Cutler, 691–698. Psychology Press.

Vennemann, Theo. 1988. *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin*. Berlin: Mouton de Gruyter.