Gradient Lexical Frequency Reflexes of the Syllable Contact Law.

Kevin B. McGowan · University of Michigan · clunis@umich.edu

The Syllable Contact Law (SCL) (Murray and Vennemann (1983), Vennemann (1988)) makes two distinct claims about the coda consonant α and the onset consonant β in a syllable contact pair: $\alpha.\beta$. First, a pair is better if the sonority of the consonant α is greater than the sonority of the consonant β. Second, the greater this sonority drop from α to β the more preferred the syllable contact pair will be[1].

It is our assumption (following, e.g. Stefan A. Frisch and Broe (2004) and Coetzee and Pater (2008)) that phonological constraints have a greater synchronic impact than merely determine the binary decision between grammatical and ungrammatical forms. Even within the set of the grammatical forms, grammar has a gradient influence, which should be realized in terms of frequency. Given this assumption, a phonological process like the SCL should have gradient frequency reflexes in the lexicon. The SCL makes a pair of straight-forward, empirically testable predictions. First, if the SCL is operating in a language one would expect that the probability of any onset β will differ from the conditional probability of β given a preceding coda α (e.g. $P(\beta|\alpha) \neq P(\beta)$). Second, to say that the SCL is underlyingly driven by the sonority hierarchy (as in e.g. Gouskova (2004)) is to predict rankings for those probabilities. Specifically, the probability of an $\alpha.\beta$ contact pair, taken as an indicator of preference, should be proportional to the sharpness of sonority drop from α to β.

This talk describes an in-progress exploration of these individual segment and segment contact probabilities in British English (BE) using the CELEX lexical database in terms of both type and token probabilities. Table 1 presents an overlay of observed BE syllable contact pair data over Gouskova's illustration of contact pairs arrayed along the sonority hierarchy. Many of the theoretically best contact pairs simply are not observed in the non-rhotic dialect described in the CELEX data, but the expected gradient frequency effects should still be visible in grammatical forms. First we estimate P(α) as the observed relative frequency of each transcribed consonant in coda position in CELEX and P(β) as the relative frequency of each transcribed consonant in onset position. These probabilities can be visualized using the Observed/Expected (O/E) (Frisch, 1996) ratios and Pointwise Mutual Information Church and Hanks (1989) for each consonant pair. Table 2 demonstrates this visualization for contact consonants. Next we estimate the joint probability of each possible consonant pair α & β in the CELEX data. Table 3 demonstrates the findings for each type of consonant pair. We perform a Chi square test for independence on the variables α and β and, finally, graph observed frequencies of sonority drop by contact pair against the predicted frequencies assuming independence. This last test demonstrates the extent to which specific observed contact pairs differ from the predictions of the SCL.

The data show that syllable contact pairs are, in fact, not conditionally independent; there is a significant effect of syllable contact. Additionally, the data do appear to support the second claim of the SCL that would predict a preponderance of *high.low* consonant pairs at syllable boundaries. There is also support for the assumed transitivity among these sonority relationships that should rank the observed probabilities in terms of the sonority hierachy. There is evidence that the more sonority falls between a pair of consonants in syllable contact the better that pair is, and this preference is reflected in the lexicon. We interpret this evidence from CELEX as confirmation of our hypothesis that even diachronic and cross-linguistic phonological constraints can be expected to have a gradient synchronic influence on the lexicon of a language.

[1] Murray & Vennemann actually state the SCL in terms of an *increase* in CONSONANTAL STRENGTH across syllable contact pairs, but subsequent work (e.g. Gouskova (2004)) tends to assume the definition given here.

References

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

Coetzee, A. W. and Pater, J. (2008). Weighted constraints and gradient restrictions on place co-occurrence in muna and arabic. *Natural Language and Linguistic Theory*, 26:289337.

Frisch, S. (1996). *Similarity and Frequency in Phonology*. PhD thesis, Northwestern University.

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

Murray, R. W. and Vennemann, T. (1983). Sound change and syllable structure in germanic phonology. *Language*, 59(3):514–528.

Stefan A. Frisch, J. B. P. and Broe, M. B. (2004). Similarity avoidance and the ocp. *Natural Language and Linguistic Theory*, 22:179228.

Vennemann, T. (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)*.

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: Overlay of BE observed & Gouskova's SCL predicted syllable contact scale

	w	ʃ	θ	r	ŋ	d	j	h	k	g	f	t	ʧ
coda	0	0.15	0.43	0	2	0.28	0	0	1.35	0.9	0.44	0.21	0.32
onset	2	1.85	1.57	2	0	1.72	2	2	0.65	1.10	1.56	1.79	1.68
	n	m	v	ʤ	s	l	p	b	ð	z			
coda	1.81	1.61	0.07	0.06	0.45	1.51	0.62	0.64	2	0.98			
onset	0.19	0.39	1.93	1.94	1.55	0.49	1.38	1.36	0	1.02			

Table 2: Observed/Expected Ratios: Contact Consonants (by type)

Pair	%	n	Fall	Pair	%	n	Fall	Pair	%	n	Fall
n-t	21.603	593	-4	s-s	1.020	28	0	z-d	0.328	9	-1
n-d	19.126	525	-2	z-n	0.947	26	1	s-d	0.328	9	1
n-s	10.747	295	-3	t-l	0.911	25	5	n-w	0.328	9	3
t-s	9.581	263	1	l-n	0.729	20	-1	s-l	0.255	7	4
t-t	6.995	192	0	d-l	0.729	20	3	s-r	0.182	5	5
s-t	3.716	102	-1	d-d	0.729	20	0	s-w	0.146	4	6
l-t	3.169	87	-5	n-n	0.692	19	0	n-r	0.146	4	2
d-n	2.732	75	2	l-z	0.656	18	-2	z-z	0.109	3	0
n-z	2.587	71	-1	s-n	0.619	17	3	t-r	0.109	3	6
l-d	1.821	50	-3	l-w	0.619	17	2	l-r	0.109	3	1
l-s	1.603	44	-4	d-t	0.619	17	-2	d-w	0.109	3	5
d-z	1.566	43	1	t-d	0.510	14	2	z-w	0.073	2	4
d-s	1.457	40	-1	n-l	0.510	14	1	t-w	0.073	2	7
t-n	1.275	35	4	z-l	0.437	12	2				

Table 3: Consonant pair types ranked by frequency