Phonological constraints in speech processing*

René Kager and Keren Shatzman
Utrecht University/UIL-OTS

1. Introduction

A growing body of literature following Pierrehumbert (1994) addresses the notion of gradient phonological well-formedness. Gradience is the phenomenon that native speakers' intuitions about the well-formedness of forms fall on a scale with multiple values, rather than into a binary division of well-formed versus ill-formed. Perhaps the best studied cases of gradient well-formedness (GWF) are those related to OCP-Place, the constraint against adjacent consonants with identical place of articulation (with adjacency being defined on the consonant projection). Gradient OCP-Place effects occur in languages as diverse as Arabic (Greenberg 1950; McCarthy 1986; Frisch, Pierrehumbert & Broe 2004), Hebrew (Berent & Shimron 1997, 2003), English (Berkley 2000), Japanese (Kawahara, Odo & Suno 2005), and Muna (Coetzee & Pater 2006).

Although GWF judgements have been shown to correlate with lexical factors such as neighbourhood density (Ohala & Ohala 1986; Bailey & Hahn 2001) and with low-level phonotactic factors such as onset and rhyme frequency (Coleman & Pierrehumbert 1997; Bailey & Hahn 2001), many studies assume an additional involvement of grammatical constraints. Paradoxically, GWF constraints are often motivated on the basis of lexical evidence alone (e.g. Frisch, Pierrehumbert & Broe 2004; Coetzee & Pater 2006). That is, the relative infrequency of forms that violate a constraint is interpreted as direct evidence for the constraint. This raises the issue to what extent the constraint has psychological significance beyond the lexical data.

Several studies support the mental reality of GWF constraints by means of native well-formedness judgements on non-words (Berent & Shimron 1997, 2003; Frisch & Zawaydeh 2001; Coetzee to appear). Still, eliciting well-formedness judgements involves a meta-linguistic task, allowing semi-conscious processes, which might be absent during speech processing, to influence listeners' judgements. This problem may be aggravated when no time limits are imposed on subjects' responses.

Hence, stronger evidence for the mental reality of GWF constraints would require an online task, eliciting subjects' responses to non-words under more realistic conditions of speech processing (Coetzee 2005; van de Weijer 2005)

Linguistics in the Netherlands 2007, **99–111**.
ISSN 0929–7332 / E-ISSN 1569–9919

and establishing correlations between response latencies and the degree of well-formedness of non-words, as predicted by phonological constraints.

Here we address the role of GWF constraints in speech processing in Dutch, using the online task of lexical decision. In this task, subjects classify stimuli as either words or non-words, and their response latencies are measured. The task is known to be strongly influenced by lexical processing factors. Hence, if an independent effect of constraints were to emerge, this would constitute strong evidence in favour of the psychological reality of those constraints. The structure of this paper is as follows. Lexical statistics on avoidance of identical place will be presented in Section 2, while Section 3 will propose three candidate constraints that potentially regulate labial co-occurrence: classical OCP-Lab, self-conjoined *Lab^2, and alignment Align-Lab. In an experimental study, reported in Sections 4–6, we will address the issue to what extent phonological constraints influence speech processing in a way that is independent of lexical factors. We will also compare the relative merits of the proposed constraints for speech processing.

2. Avoidance of identical place in the Dutch lexicon

Dutch shows an under-representation of stems in which neighbouring consonants share place of articulation. Lexical statistics were obtained from the CELEX database (Baayen, Piepenbrock & Gulikers 1995). We first created a lexicon of underived stems, containing 8,305 items, and a total of 40,517 phoneme tokens. To assess under-representation of CVC sequences with consonants of identical place of articulation, we calculated Observed/Expected ratios (Frisch 1996). This ratio divides the observed number of C_1VC_2 sequences by the number expected when C_1 and C_2 were to combine freely. The expected number equals

$$p(C_1) * p(C_2) * N(CVC)$$

Here, $p(C_1)$ is the probability that C_1 occurs initially in a CVC sequence; $p(C_2)$ the probability that C_2 occurs finally in a CVC sequence; and N(CVC) the total number of CVC tokens in the lexicon (N = 11,092). O/E ratios for natural classes, such as PVP (labial-vowel-labial), were calculated by summing observed and expected numbers over all individual CVC sequences, and taking the ratio of the sums. Labials were defined as the set /p, b, f, v, m/, coronals as /t, d, s, z, n/, and dorsals as /k, g, x, ɣ, ŋ/.

Results are as follows. For stem-initial CVC sequences, there is no evidence of under-representation for strictly identical pairs of labials and dorsals. Under-representation is severe (with O/E ratios falling below 0.5) between pairs of non-identical labials and dorsals. Coronal pairs are only slightly under-represented.

Table 1. O/E ratios for initial CVC sequences sharing place of articulation (P: labial, K: dorsal, T: coronal) in in a lexicon of 8,305 Dutch stems.

	PVP	KVK	TVT
Identical pairs	0.956	0.913	0.725
Non-identical pairs	0.444	0.436	0.880

Turning to non-initial CVC sequences, we find that under-representation is severe across the board except for non-identical coronal pairs.

Table 2. O/E ratios for non-initial CVC sequences sharing place of articulation (P: labial, K: dorsal, T: coronal) in a lexicon of 8,305 Dutch stems.

	PVP	KVK	TVT
Identical pairs	0.134	0.218	0.281
Non-identical pairs	0.091	0.347	0.827

With increased distance between the members of consonant pairs, the effects of similarity avoidance vanish. Counts are based on initial $C_1VC_2VC_3$ sequences, comparing consonant pairs that are adjacent across an intervening vowel (C_1VC_2 and C_2VC_3) with non-adjacent pairs (C_1VXVC_3). No distinction is made between identical and non-identical pairs.

Table 3. O/E ratios for pairs of adjacent and non-adjacent consonants sharing place of articulation (P: labial, K: dorsal, T: coronal) in initial CVCVC sequences in a lexicon of 8,305 Dutch stems.

	P…P	K…K	T…T
# C_1 V C_2 V X	0.529	0.600	0.785
# X V C_2 V C_3	0.319	0.406	0.873
# C_1 V X V C_3	1.121	1.193	0.982

Summarizing the lexical distribution, we find that homorganic labial and dorsal pairs are severely under-represented, except in identical pairs in initial position. Moreover, no under-representation occurs for non-adjacent pairs.

3. Constraints

To establish whether lexical under-representation is *grammatically* represented by a GWF constraint, candidate constraints are needed which might capture the lexical distributional patterns. We focus on constraints on the distribution of labials, which is the focus of our experimental study. Clear choices are two anti-co-occurrence constraints:

(1) OCP-Lab No adjacent labials on the consonantal tier.

(2) *Lab^2 No two labials per word.

The difference between classical OCP-Lab (McCarthy 1986) and self-conjoined *Lab^2 (Alderete 1997, Itô & Mester 2003) is whether adjacency matters ('yes' for OCP-Lab, 'no' for *Lab^2). Our lexicon study suggests that classical OCP should be more relevant than *Lab^2 since we found that adjacent CVC sequences are more severely under-represented than non-adjacent CVC sequences.

Yet co-occurrence constraints fail to capture the relevance of edges to under-representation: we learned that non-initial PVP sequences are more severely under-represented than initial ones. In Optimality Theory, reference to edges is captured by alignment. Hence, we tentatively introduce Align-Lab:

(3) Align-Lab Every labial must be word-initial.

Align-Lab adds one violation for each labial which is non-initial. A non-initial PVP sequence incurs two violation marks, while an initial PVP sequence incurs one. Note that Align-Lab is not a co-occurrence constraint. Multiple violation in non-initial labials simulates some of the effects of OCP-Lab and *Lab^2, and hence the violation patterns of the three constraints will overlap to a large extent.

Labial alignment is supported by lexical statistics: 45.4% of all labials are initial, compared to 29.9% of coronals. A chi-square test indicated that, compared to coronals, labials are significantly over-represented in initial position, c2 (1, N = 13,354) = 329.87, $p < .0001$.

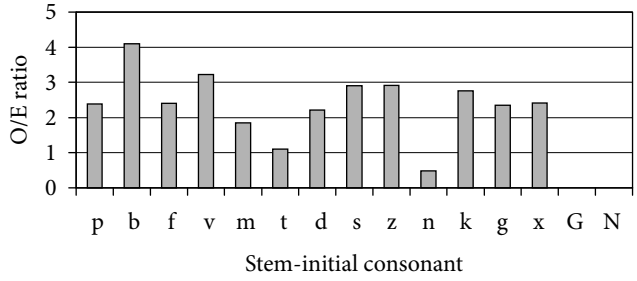

Figure 1. O/E ratios for stem-initial consonants in a lexicon of 8,305 Dutch stems

The general motivation for featural alignment comes from long-distance assimilation, specifically vowel harmony (Kirchner 1993, Cole & Kisseberth 1994, Akinlabi 1996; cf. Piggott 2000). Fikkert & Levelt (2002/6) have proposed a constraint [Labial "Word starts with labial" on the basis of Dutch children's early productions, featuring a pattern of consonant harmony conspiring towards initial labials. Yet the authors state reservations as to whether this constraint survives into adult

grammar. Finally, Fikkert (2007) presents experimental evidence for a labial-initial preference in Dutch children.

To bring out the differences between the three constraints more clearly, an overview of violations is offered below for a range of non-words. Note that this is not a tableau.

Table 4. Violation patterns for three constraints.

CVCVC	OCP-Lab	*Lab^2	Align-Lab
pamap	**	**	**
tamap	*	*	**
pamat	*	*	*
panap		*	*
tamat			*
panat			
sCVCVC	**OCP-Lab**	***Lab^2**	**Align-Lab**
spamap	**	**	***
stamap	*	*	**
spamat	*	*	**
spanap		*	**
stamat			*
spanat			*

Three crucial differences occur between constraints. Labial co-occurrence (for example, *pamat* vs. *tamat*) matters to OCP-Lab and *Lab^2, not to Align-Lab. Distance between labials (for example, *pamat* vs. *panap*) matters only to OCP-Lab, not to Align-Lab or *Lab^2. Strict initiality (for example, *pamat* vs. *tamap*) matters only to Align-Lab, not to OCP-Lab or *Lab^2. No constraint captures all three determinants of under-representation: the distance between labials, their initiality, and co-occurrence.

We emphasize that the internalization of constraints is ultimately affected by learnability considerations. Because of the non-local interaction between pairs of labials (across a vowel for OCP-Lab, and unbounded for *Lab^2), attributing a lexical gap to a particular constraint involves quantitative analysis of long spans of segments. In general, non-local co-occurrence restrictions are more difficult to detect for learners than local ones, although dependencies between non-adjacent segments are easier to learn than dependencies between non-adjacent syllables (Newport & Aslin 2004). Moreover, Pierrehumbert (2003) argues that the longer the segmental span (or *n*-phone) to be analyzed, the more difficult it becomes for learners to detect under-representation. This is because in a lexicon of finite size, the observed and expected numbers for long *n*-phones quickly fall below reliable values. Hence, the detection of a 'long' constraint, one spanning a triphone C_iVC_i

(in case of OCP-Lab) or unbounded distance $C_i \ldots C_i$ (in case of *Lab^2), should be less robust than the detection of a 'short' constraint (Align-Lab). Assuming categorical alignment (McCarthy 2003), learners can spot misalignment between a labial and the left stem edge within a biphone-sized window.

4. Purposes of the experimental study

As we stated earlier, lexical under-representation alone fails to prove the psychological reality of a GWF constraint. That is, under-representation may be accidental, rather than internalized. Also, the cause of under-representation may reside in production, without affecting perception. We maintain that the psychological reality of a constraint and its involvement in speech processing should be confirmed by demonstrating its independent effect on listeners' perception of speech. We chose to submit constraints to a particularly strong test, one involving speech processing.

The general question we address is: Do GWF constraints influence speech processing? We selected an online task in which subjects respond to non-words which violate or satisfy the constraints in various degrees, and in which reaction times could reflect gradient effects of constraints. To establish the psychological reality of GWF constraints, it needs to be demonstrated that their effects are truly independent of general factors which are known to influence the processing of words and non-words, in particular lexical neighbourhoods and cohorts, as well as low-level phonotactic factors such as transitional probabilities (Vitevitch & Luce 1999). A demonstration of such an independent effect of constraints would indicate that phonological knowledge is not exclusively lexical in nature, adding plausibility to the idea of a phonological grammar. The lexical decision task allows us to examine whether constraints influence speech processing independently of general processing factors. Due to the lexical orientation of the task, relatively few lexical decision studies have used non-words. With non-words, lexical factors are still dominant, but there is an increased role of low-level phonotactic probability. However, the role of gradient abstract phonological constraints has thus far not been tested.

In case the first question is answered positively, a follow-up emerges: Which constraint (OCP-Lab, *Lab^2, or Align-Lab) is the best predictor of subjects' responses? Evidently, labial co-occurrence constraints are supported by a wealth of cross-linguistic evidence (McCarthy 1986; Odden 1988; Yip 1988), while a range of experimental studies have established the psychological reality of OCP-Lab, in tasks such as word-likeness judgements (for Hebrew: Berent & Shimron 1997, 2003; Arabic: Frisch & Zawaydeh 2001; English: Coetzee to appear) and phoneme

identification (English: Coetzee 2005). Yet, as shown, the relevance of edges to labial under-representation in Dutch is not captured by co-occurrence constraints; hence, we give Align-Lab a chance to prove its relevance.

We set out to test the relevance of constraints using an online processing task. Questions to be addressed in the experimental study are the following. First, do GWF constraints influence speech processing, specifically response latencies to non-words in lexical decision? Second, in case this turns out correct, which constraint (OCP-Lab, *Lab^2, or Align-Lab) predicts response latencies best?

Predictions are as follows. First, we expect reaction times to be influenced by lexical factors. That is, non-word stimuli should be classified as non-words more slowly if they carry more resemblance to existing words. Second, if there is a constraint effect, reaction times to non-word stimuli should be influenced by their degree of phonotactic well-formedness, as is assessed by GWF constraints. Thus, non-words with more serious violations of constraints should show shorter reaction times than non-words with less serious violations.

5. Method

Participants were 20 students of Utrecht University, all native speakers of Dutch.

Stimuli were 96 disyllabic non-word sequences $C_1VC_2VC_3$ and $sC_1VC_2VC_3$ in which labials /p,f,m/ and coronals /t,s,n/ occupied the three consonant slots. This produced eight combination types: PPP, PPT, PTP, PTT, TPP, TPT, TTP, TTT, each type occurring without and with initial /s/. For each combination type, there were six manner patterns: plosive-nasal-plosive, plosive-nasal-fricative, plosive-fricative-nasal, nasal-fricative-nasal, nasal-plosive-fricative, and nasal-plosive-nasal. Each of 96 types occurred twice, using different vowel melodies, giving a total of 192 non-word stimuli. Stress was uniformly final.

It turned out to be difficult to avoid differences in the lexical properties (i.e., the degree of similarity to real words) of the various stimulus types. Hence, we decided to tease apart the effects of phonological constraints from the effects of lexical factors by a multi-level regression analysis. The following lexical factors were considered: Lexical Neighbourhood Density (LND), Transitional Probabilities, Cohort Density, and Isolation Point.

To the 192 non-word test items, we added 192 words and 40 non-word distractors, giving a total of 424 stimuli. These were read by a female speaker who was naive of the purposes of the study. The stimuli, recorded digitally, were played to subjects over headphones in a soundproof booth. Responses to non-words were made by the subjects' dominant hand.

6. Results

The mean reaction times (RT) per stimulus type are shown in Figure 1. There was considerable variation between the shortest average RTs (around 870 msecs for PPP) and the longest (around 990 msecs for sTTT). RTs to sCVCVC items were generally longer than RTs to CVCVC items, suggesting a stimulus duration effect. However, the variation between stimulus types fails to prove our case: variation in RTs need not be due to constraints, but could be entirely due to lexical factors.

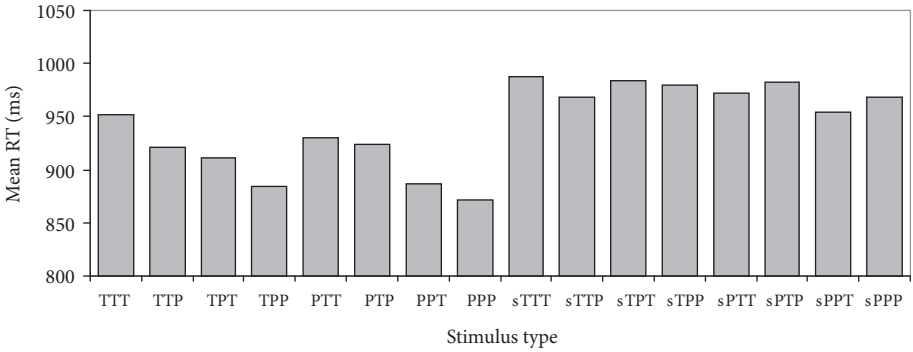

Figure 2. Mean reaction time (in milliseconds) per stimulus type

To prove the relevance of constraints, we must establish that their effects are independent of the lexicon. This can be shown if their effects are teased apart from general processing factors. Hence, we need a statistical model to assess the effects of constraints independently from the effects of other factors. A linear regression model can identify the factors that explain the variation in RTs. We will refer to these as *predictors*. RTs in lexical decision are partly a function of the activation and competition of words in the lexicon. We chose three lexical predictors which are likely to affect RTs: Lexical Neighbourhood Density, Cohort Density, and Isolation Point. In addition, Biphone Transitional Probability was included as a predictor assessing the influence of low-level phonotactic probability. Stimulus Duration and Stimulus Length (the number of segments: 5 for CVCVC, 6 for sCVCVC) were included as predictors in order to control for the influence of stimulus length on RTs.

These six predictors were combined in a linear regression model, which we will refer to as 'model A', with RT as the dependent variable. RT was averaged per item across subjects, giving a total of 192 data points. The effect of each predictor was examined given that all the other predictors are already in the model. It turned out that only two predictors were significant: Cohort Density (F (1,185) = 24.58, $p < 0.0001$) and Stimulus Duration (F (1, 185) = 30.96, $p < 0.0001$). Significance of

Cohort Density shows the relevance of lexical processing: RTs were longer when there were many existing words overlapping with the initial phonemes of the non-word.

The second step involved the addition of phonological constraints to the initial model, one at a time: OCP-Lab, *Lab^2, and Align-Lab. Again, the effect of each predictor was examined given that all the other predictors are in the model. Thus the independent effect of each constraint was tested by examining whether adding this constraint to the model significantly increased the explained variance in RTs. It is the increase in explained variance following the addition of a constraint that establishes its independence from the other predictors. We found that each constraint, when added individually to the model, turned out to have a small but significant effect.

Table 5. ANOVA results for the effects of the phonotactic constraints

Model	Predictor	ANOVA	Gained variance
Model A +	OCP-Lab	$F = 5.41$, $p < 0.05$	3.2%
Model A +	*Lab^2	$F = 4.11$, $p < 0.05$	2.3%
Model A +	Align-Lab	$F = 8.18$, $p < 0.005$	4.7%

Note. Degrees of freedom are 1, 184

This regression analysis provides a clear answer to our first question: GWF constraints indeed influence speech processing, independently of lexical factors.

This brings us to our second question: which of the three constraints is the best predictor? The fact that each constraint taken individually is a significant predictor (as shown above) does not imply that each constraint contributes equally to explaining the variance in RTs. The effects of constraints are highly confounded (i.e., many predictions are shared by all three constraints). To compare the effects of the constraints with each other, we entered all three constraints into model A, and examined which constraint still had a significant effect when the other constraints were already in the model. In this analysis, therefore, the effect of each constraint was evaluated after partialling out the effects of the other variables, including the two other constraints.

As before, there was a robust and significant inhibitory effect of Cohort Density (F (1, 182) = 23.38, $p < 0.0001$), and Stimulus Duration (F (1, 182) = 29.53, $p < 0.0001$), with no other significant effects of lexical factors or of stimulus length. With respect to the constraints, the results showed that after entering *Lab^2 and Align-Lab to model A, OCP-Lab did not have a significant effect ($F < 1$). Similarly, when Align-Lab and OCP-Lab were included in model A, *Lab^2 had no additional significant effect (F (1, 182) = 1.02, $p > 0.1$). However, when entering Align-Lab to the model with OCP-Lab and *Lab^2 already included, the effect of Align-Lab still emerged significant (F (1, 182) = 3.72, $p = 0.05$). This result

indicates that after partialling out the effects of all the other factors, there is still a significant effect of Align-Lab on response latencies. In contrast, when the effect of Align-Lab is first partialled out, the effects both OCP-Lab and *Lab^2 are no longer significant. This suggests that the variance in the data that is explained by the phonotactic factors is best captured by the constraint Align-Lab.

7. Discussion

The results of the lexical decision study suggest a role for GWF constraints in speech processing. The effects of the constraints, though modest, were independent of major lexical predictors, such as Cohort Density. This suggests that the three constraints under investigation are relatively independent of the lexicon. Moreover, the effects of the constraints were independent of lower-level phonotactics, such as transitional probabilities. This suggests that GWF constraints are relatively abstract. Of the three constraints put to the test, Align-Lab turned out to be the best predictor of RTs, as compared to co-occurrence constraints OCP-Lab and *Lab^2. Arguably, Align-Lab is more robustly learnable from finite input since it is biphone-based (while co-occurrence constraints span longer distances). GWF constraints should ultimately be evaluated on the criterion how well they predict experimental data in speech processing, and not just on how well they match lexical frequencies.

Notes

* This research was supported by a grant from the Netherlands Organization for Scientific Research (NWO) (277–70–001).

1. See, among others, Frisch (1996); Berent & Shimron (1997); Coleman & Pierrehumbert (1997); Berkley (2000); Hayes (2000); Zuraw (2000); Bailey & Hahn (2001); Frisch & Zawaydeh (2001); Boersma & Hayes 2001 (2001); Berent & Shimron (2003); Hammond (2003); Frisch, Pierrehumbert & Broe (2004); Pater (2005); Coetzee (2005); Coetzee & Pater (2006); Hayes & Londe (2006); Hayes & Wilson (2006); Berent, Steriade, Lennertz & Vaknin (2007); Coetzee to (appear).

2. The stem lexicon contained a phonological transcription of all lemmas marked as morphologically underived in CELEX DML. All homophones were removed.

3. Omitting liquids, which behave as a-typical coronals in Dutch and other languages.

4. CELEX contrasts voicing in dorsal fricatives.

5. No evidence was found that OCP effects are stronger in tauto-syllabic CVC sequences. Differences were found between initial tauto-syllabic and hetero-syllabic CVC sequences. For example, tauto-syllabic PVP sequences display a smaller difference between identical pairs (0.743) and non-identical pairs (0.530) than hetero-syllabic PVP sequences do (1.121 versus 0.376, respectively).

6. Perceptual explanations for OCP effects have been offered by Boersma (1998, 2000) and Frisch (2004).

7. Closer analysis of word-likeness judgements has revealed an independence between lexical factors and low-level phonotactics (Vitevitch & Luce 1999; Bailey & Hahn 2001).

8. Coetzee's study focussed on /sCVC/ non-words with strictly identical consonants.

9. Mulder (in progress) used our set of stimuli in a comparative word-likeness judgement task (Berent & Shimron 1997; Coetzee to appear), and found results very similar to ours.

10. A lexical neighbour is any word that arises by either changing, inserting, or deleting a phoneme in a non-word. Lexical Neighbourhood Density was calculated as the sum of logged frequencies of an item's lexical neighbours.

11. The probability that a segment occurs given a preceding segment. For a given non-word, this was calculated as the logged product of its biphone transitional probabilities, where word boundary symbols were counted as segments.

12. A cohort is a set of words which share a sequence of segments at their beginning, such as /kɔf/: *koffie, koffer*. The members of a cohort are in lexical competition. Cohort density was calculated as the sum of token frequencies of all members of a cohort.

13. The point at which a non-word's cohort runs empty, and where it can be decided to be a non-word. This was counted by the number of segments preceding the isolation point.

References

Akinlabi, Akinbiyi. 1996. "Featural affixation". *Journal of Linguistics* 32.239–289.

Alderete, John. 1997. "Dissimilation as local conjunction". *Proceedings of North East Linguistics Society* 27.17–32.

Baayen, Harald, Richard Piepenbrock & Leon Gulikers. 1995. *The CELEX Lexical Database*. Philadelphia: Linguistics Data Consortium, University of Pennsylvania.

Bailey, Todd & Ulrike Hahn. 2001. "Determinants of wordlikeness: Phonotactics or lexical neighborhoods?" *Journal of Memory and Language* 44.568–591.

Berent, Iris & Joseph Shimron. 1997. "The representation of Hebrew words: Evidence from the Obligatory Contour Principle". *Cognition* 64.39–72.

Berent, Iris & Joseph Shimron. 2003. "Co-occurrence restrictions on identical consonants in the Hebrew lexicon: Are they due to similarity?" *Journal of Linguistics* 39.31–55.

Berent, Iris, Donca Steriade, Tracy Lennertz & Vered Vaknin. 2007. "What we know about what we have never heard: Evidence from perceptual illusions". *Cognition* 104. 591–630.

Berkley, Deborah M. 2000. Gradient OCP effects. PhD diss., Northwestern University.

Boersma, Paul. 1998. "The OCP in functional phonology". *Proceedings of the Amsterdam Institute of Phonetic Sciences* 22.21–45.

Boersma, Paul. 2000. The OCP in perception grammar. Ms., University of Amsterdam. [Rutgers Optimality Archive 435.]

Boersma, Paul, & Bruce Hayes. 2001. "Empirical tests of the Gradual Learning Algorithm". *Linguistic Inquiry* 32.45–86.

Coetzee, Andries. 2004. What it means to be a loser: Non-optimal candidates in optimality theory. PhD diss., University of Massachusetts.

Coetzee, Andries. 2005. "The OCP in the perception of English". *Prosodies* ed. by S. Frota, M. Vigario & M.J. Freitas, 223–245. New York: Mouton de Gruyter.

Coetzee, Andries. to appear. "Grammar is both categorical and gradient". *Phonological Argumentation*, ed. by S. Parker. London: Equinox Publishers.

Coetzee, Andries & Joe Pater. 2006. Lexically Ranked OCP-Place Constraints in Muna. Ms., University of Michigan and University of Massachusetts, Amherst.

Cole, Jennifer & Charles Kisseberth. 1994. "An Optimal Domains theory of harmony". *Studies in the Linguistic Sciences*. 24.101–114.

Coleman, John & Janet Pierrehumbert. 1997. "Stochastic phonological grammars and acceptability". *Computational Phonology: Third meeting of the ACL special interest group in computational phonology*, 49–56. Association for Computational Linguistics.

Fikkert, Paula. 2007. Constraints on place of articulation. Paper presented at Linguistics in the Netherlands, Utrecht, 3 February, 2007.

Fikkert, Paula, & Claartje Levelt. 2002/2006. How does Place fall into Place? The lexicon and emergent constraints in children's developing phonological grammar. Ms., Radboud University of Nijmegen & Leiden University.

Frisch, Stefan. 1996. Similarity and frequency in phonology. PhD diss., Northwestern University.

Frisch, Stefan. 2004. "Language processing and segmental OCP effects". *Phonetically-Based Phonology* ed. by B. Hayes, R. Kirchner & D. Steriade, 346–371. Cambridge: Cambridge University Press.

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

Frisch, Stefan & Bushra Zawaydeh. 2001. "The psychological reality of OCP-place in Arabic". *Language* 77.91–106.

Greenberg, Joseph. 1950. "The patterning of root morphemes in Semitic". *Word* 5.62–181.

Hammond, Michael. 2003. "Phonotactics and probabilistic ranking". *Formal Approaches to Function in Grammar: In Honor of Eloise Jelinek* ed. by A. Carnie, H. Harley & M. Willie, 319–332. Benjamins, Amsterdam.

Hayes, Bruce. 2000. "Gradient well-formedness in optimality theory". *Optimality Theory: Phonology, Syntax, and Acquisition* ed. by J. Dekkers, F. van der Leeuw & J. van de Weijer, 88–120. Oxford: Oxford University Press.

Hayes, Bruce & Zsuzsa Londe. 2006. "Stochastic phonological knowledge: The case of Hungarian vowel harmony". *Phonology* 23.59–104.

Hayes, Bruce & Colin Wilson. 2006. A maximum entropy model of phonotactics and phonotactic learning. Ms., University of California, Los Angeles.

Itô, Junko & R. Armin Mester. 2003. *Japanese Morphophonemics: Markedness and Word Structure*. Boston: MIT Press.

Kawahara, Shigeto, Hajime Ono & Kiyoshi Sudo. 2006. "Consonant co-occurrence restrictions in Yamato Japanese". *Japanese/Korean Linguistics* 14.27–38.

Kirchner, Robert. 1993. Turkish vowel harmony and disharmony: an Optimality Theoretic account. Paper presented at Rutgers Optimality Workshop.ROW-I. New Brunswick, N.J. October 23, 1993.

McCarthy, John. 1986. "OCP effects: gemination and anti-gemination". *Linguistic Inquiry* 17.207–263.

McCarthy, John. 2003. "OT constraints are categorical". *Phonology* 20.75–138.

Mulder, Merthe. in progress. Labial co-occurrence constraints in Dutch gradient well-formedness. Master's thesis, Utrecht University.

Newport, Elissa & Richard Aslin, 2004. "Learning at a distance: I. Statistical learning of non-adjacent dependencies". *Cognitive Psychology* 48.127–162.

Odden, David. 1988. "Anti anti-gemination and the OCP". *Linguistic Inquiry* 19.451–475.

Ohala, J.J. & M. Ohala. 1986. "Testing hypotheses regarding the psychological manifestation of morpheme structure constraints". *Experimental phonology*, ed. by J.J. Ohala & J. J. Jaeger, 239–252. Orlando, FL: Academic Press.

Pater, Joe. 2005. "Learning a stratified grammar". *Proceedings of the 29th Boston University Conference on Language Development* ed. by A. Brugos, M.R. Clark-Cotton, & S. Ha, 482–492. Somerville, MA: Cascadilla Press.

Pierrehumbert, Janet. 1994. "Syllable structure and word structure: A study of triconsonantal clusters in English". *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology* III ed. by P.A. Keating, 168–188. Cambridge: Cambridge University Press.

Pierrehumbert, Janet. 2003. "Probabilistic phonology: Discrimination and robustness". *Probability Theory in Linguistics* ed. by R. Bod, J. Hay & S. Jannedy, 177–228. Cambridge, MA: MIT Press.

Piggott, Glyne. 2000. "Against featural alignment". *Journal of Linguistics* 36.85–129.

Vitevitch, Michael & Paul Luce. 1999. "Probabilistic phonotactics and spoken word recognition". *Journal of Memory & Language* 40.374–408.

van de Weijer, Joost. 2005. "Listeners' sensitivity to consonant variation within words". *Lund University, Dept. of Linguistics Working Papers* 51.225–238.

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

Zuraw, Kie. 2000. Patterned exceptions in phonology. PhD diss., University of California, Los Angeles.