STEFAN A. FRISCH, JANET B. PIERREHUMBERT and MICHAEL B. BROE

SIMILARITY AVOIDANCE AND THE OCP*

ABSTRACT. It has long been known that verbal roots containing homorganic consonant pairs are rare in Arabic, motivating the existence of an OCP-Place constraint (Obligatory Contour Principle on place of articulation) in the phonological grammar. We explore this constraint using an on-line lexicon of Arabic roots. The strength of the constraint is quantified by the ratio of the observed number of examples of each consonant pair to the number that would be statistically expected under random combination of phonemes. We show that the strength of the effect over all pairs is a gradient function of the similarity of the consonants in the pair. A similarity metric based on natural classes is developed, which solves the formal difficulties of contrastive underspecification theory while preserving the insight that contrastiveness plays an important role in perceived similarity. This metric is applied in an explicit model of the gradient OCP constraint, which achieves a better fit to the regularities and sub-regularities of the Arabic verbal lexicon than any prior approach. Lastly, we review evidence for the psychological reality of the constraint, for its existence in related forms in other languages, and for its cognitive/phonetic foundations in the speech processing system. We argue that the total body of evidence supports a model in which phonetic and cognitive pressures incrementally affect the lexicon, and phonotactic constraints are abstractions over the lexicon of phonological forms.

0. INTRODUCTION

In this paper, we present a new account of the *OCP-Place* constraint in the verbal roots of Arabic. OCP-Place (or the Obligatory Contour Principle for Place of Articulation) refers to a phonotactic constraint that disfavors combinations of homorganic consonants in proximity to each other (McCarthy 1986, 1988, 1994). The highly elaborated consonantal system

* Parts of this manuscript have previously been circulated as Frisch et al. (1995) and (1997) on the Rutgers Optimality Archive. We would like to thank the many people who provided comments on this early work, including Joan Bybee, Grzegorz Dogil, Ken de Jong, Dan Jurafsky, Michael Kenstowicz, Peggy MacEachern, Barry Schein, Dan Silverman, Richard Wright, several anonymous reviewers, and the many audiences who have heard parts of this work at conferences and colloquia.

This work was supported in part by NSF Grant No. BNS-9022484 to Northwestern University, NIH Training Grant No. DC00012 to Indiana University, a fellowship from the John Simon Guggenheim Memorial Foundation to Janet Pierrehumbert, and the *Language Learning* Research Assistant Professorship at the University of Michigan.

Natural Language & Linguistic Theory **22:** 179–228, 2004.

of Arabic provides an opportunity to explore in detail the nature of the OCP-Place constraint. The analysis of consonant co-occurrence patterns in Arabic reveals two fundamental insights. First, the co-occurrence constraints are gradient and their robustness can be related to a scalar measure of similarity between homorganic consonant pairs. Second, the similarity metric is sensitive to the contrastiveness of phonological features, but is also faithful to the surface realization of segments.

We explain the co-occurrence patterns in the Arabic roots as a diachronic result of a processing constraint that disfavors repetition. We refer to this constraint as *similarity avoidance*. We claim that consonant sequences that are difficult to process are less likely to exist because they are at a competitive disadvantage in comparison to alternatives that are easier to process. There is also evidence that Arabic speakers are implicitly aware of the OCP-Place constraint as a gradient constraint (Frisch and Zawaydeh 2001, discussed in section 4.1). We believe that this evidence supports a theory of phonological knowledge in which phonetic and lexical information are the foundation upon which the segment inventory and phonotactic constraints are built. In our account, the functional diachronic origin of similarity avoidance in the lexical patterns of Arabic is distinct from the formal synchronic psychological representation of similarity avoidance as OCP-Place in the minds of speakers. However, they are strongly related: The phonotactic patterns that can be observed in the lexicon directly determine the mental representation of the phonotactic constraints. This account contrasts with previous treatments of consonant co-occurrence in Arabic that assume the constraints involved are formal symbolic statements selected from a universal inventory of possible constraints. In positing the OCP as a universal building block of grammars, previous accounts claim that the existence of the OCP-Place pattern in Arabic follows only from the availability of this constraint as an option in universal grammar. The most comprehensive previous treatment is a series of papers by McCarthy (1986, 1988, 1994). The present treatment is an extension of the similarity account provided in Pierrehumbert (1993).

In describing and explaining the phonotactic patterns in any language, phonologists recognize a distinction between accidental and systematic gaps in the lexicon. In practice, this distinction has often been applied in an intuitive and *post-hoc* fashion. A gap is taken to be systematic if it belongs to a natural class of examples according to the current theory of the investigator. Otherwise it is viewed as accidental. Following Greenberg (1950) and McCarthy (1988, 1994), we take a more data-driven approach to discovering phonotactic constraints that compares the existing verbal roots of Arabic to the set of roots that would be expected if there were no phonological constraints over the roots. In other words, the observed counts of consonantal combinations in the lexicon are compared to the counts that would be expected under the null hypothesis that the consonants can freely combine at random. It is known that speakers are implicitly aware of statistical patterns in the lexicon (Greenberg and Jenkins 1964; Vitevitch et al. 1997; Coleman and Pierrehumbert 1998; Vitevitch et al. 1999; Frisch et al. 2000; Treiman et al. 2000; Bailey and Hahn 2001). Therefore, we feel it is worthwhile to consider an account of the phonotactic constraints of a language that explains not only the individual systematic gaps that have been the focus of traditional phonological inquiry, but also the entire pattern of statistical under-representation and over-representation of segmental combinations in the language.

The theoretician can interpret the range of variation found in a statistical analysis of phonological data in two very different ways. On the one hand, the traditional goal of generative grammar has been to identify the constraints that are needed in order to properly classify any particular form as grammatical or ungrammatical. In line with this approach, we could propose that the grammar should exclude patterns whose under-representation is large. Statistical under-representation is assumed to signify cases where the pattern in question is phonologically illegal, even though a few exceptional forms may be found. This is the interpretation of under-representation (and phonological constraint) that is carried out formally in McCarthy (1988, 1994).

In an alternative perspective, developed in Pierrehumbert (1994) and increasingly used elsewhere (Anttila 1998; Broe 1997; Boersma 1998; Boersma and Hayes 2001; Buckley 1997; Frisch 2000a; Hay et al. 2003; Hayes and MacEachern 1998; Kessler and Treiman 1997; Treiman et al. 2000), a native speaker's knowledge of language has a stochastic component. Under this view, there is a model of the observed statistical patterns in the native speaker's phonology. McCarthy (1994) acknowledges this point of view and discusses the appropriateness of such a 'soft' constraint for Arabic, but does not carry the point through formally.

This paper carries through such an analysis and attempts to deal with the issues and implications that arise in broadening the scope of phonological analysis to include quantitative patterns. In a quantitative analysis of the co-occurrence patterns in the Arabic roots, we find that different patterns may be under-represented or over-represented to different degrees. However, the variation in the degree of co-occurrence among violations of OCP-Place is systematic, not random. In this paper, we differentiate between a soft constraint and a gradient constraint. A *soft constraint* is a constraint that is held to be part of the grammar even though it is sometimes violated. In Optimality Theory, any constraint that can be violated in order to meet a higher-ranked constraint is a soft constraint. In some other models, constraints can have probabilities, and a soft constraint is one whose probability is less than 1.0 (i.e., the requirements of the constraint are not met in all cases). A *gradient constraint* is a constraint that is quantitatively sensitive to violations of different degrees, such that forms that violate the constraint to a lesser degree are more frequent than forms that violate the constraint to a greater degree. Gradient constraints are soft (because they are sometimes violated), but not all soft constraints are gradient. Claiming that a constraint is gradient means that it is possible to establish a scale of softness such that the probability of a violation is a well-behaved function of some other factor. In the case of the OCP, this factor is similarity. We believe that the correct model of OCP-Place in Arabic is a gradient constraint, with the stringency of the constraint being a function of the similarity of the consonants involved. Quantitative co-occurrence patterns can thus be explained by a single phonetically-motivated constraint of similarity avoidance. Further, we believe that the quantitative knowledge of this gradient pattern is psychologically real for Arabic speakers, in the sense of being applied productively in well-formedness judgments and neologisms.

The gradient patterns of consonant co-occurrence in Arabic are analogous to categorical phonotactic patterns discussed elsewhere (e.g., MacEachern 1999; Yip 1989). We propose that the realization of similarity avoidance constraints in the world's languages falls on a continuum of strength from the gradient to the categorical. Frisch (1996), Frisch et al. (1997), and Frisch (2000a) develop a formalism that uses the logistic function to treat the degree of gradience of a phonotactic constraint in terms of the sharpness of a grammatical categorization function over a set of linguistically-related patterns. Examples of logistic functions that reflect different degrees of gradience are shown in Figure 1. Analogous cases of categorization with different degrees of gradience have previously been demonstrated in phonetic studies. For example, the nearly linear function is appropriate for the manipulation of pitch range discussed by Liberman and Pierrehumbert (1984), while the categorical function is appropriate for the effect of changes in cues on phoneme perception for consonants (Repp 1984). Most instances of phonetic categorization fall somewhere in between. We claim that gradient relations are also found in phonotactic constraints. In a gradient phonotactic constraint, knowledge of the degree of under-representation or over-representation of a form is imputed to the minds of speakers, and part of the information contained in the speaker's grammar is the observed probabilities of the various outcomes.

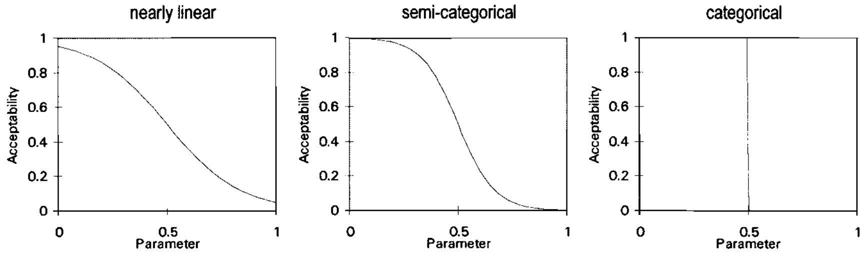

Figure 1. Three constraints with different degrees of gradience representing possible phonotactic constraints.

In short, we propose that both within Arabic and across languages, categorical and gradient instances of OCP constraints should be explained by the same mechanism. The extent to which the particular form of the constraint is categorical (i.e., the extent to which the categorization over the similarity scale resembles a step function) depends on many factors. Descriptions in the literature imply that relevant factors probably include other aspects of the language's phonological and morphological structure, interacting phonotactic constraints, and the language's history.

1. Consonant Co-Occurrence in Arabic

We begin our analysis of the Arabic verbal root morphemes with a review of the original quantitative description given by Greenberg (1950) and the autosegmental account of Arabic morphology and the OCP drawn primarily from McCarthy (1986, 1988, 1994), Mester (1986), and Padgett (1995). The consonant inventory of Arabic is given in (1).[1] Arabic verbal roots consist of a set of two to four consonants, with the canonical root containing three consonants. Vowels are inserted between the consonants to make word forms according to a CV template, an example of a non-concatenative morphological system. For example, the root /k t b/ has among its word forms *katab-a* 'he wrote', *kutib-a* 'it was written', and *kuttib-a* 'he was made to write'.

[1] The segments /$t^ʕ$, $d^ʕ$, $s^ʕ$, $z^ʕ$/ are often called 'emphatic' consonants. They are similar to the ordinary coronal consonants /t d s z/ but with an additional vocal tract constriction in the pharyngeal region. Note that this inventory is an idealized inventory that roughly corresponds to the historical basis of the modern Arabic dialects and Standard Arabic as studied by Greenberg (1950) and McCarthy (1986, 1988, 1994).

(1)

labial	dental	alveolar	palatal	velar	uvular	pharyngeal	laryngeal
b f		t d tˤ dˤ		k g	q		ʕ
	θ ð	s z sˤ zˤ	ʃ		χ ʁ	ħ ʕ	h
m		n					
		l r					
w			j				

The co-occurrence restrictions described by Greenberg are based only on the consonantal roots, and not on any derived forms. Greenberg first observes that there are no roots that repeat the same consonant in first and second position (e.g., **dadam*, though McCarthy (1994) reports one verb of this form). Many verbs are found with identical consonants in the second and third positions of the root. Examples include *madad* 'stretch' and *farar* 'flee'. Greenberg shows that, more generally, Arabic consonants divide into groups of homorganic consonants that tend not to co- occur within the same root (apart from the pairs of identical consonants in the second and third position just mentioned).[2] McCarthy (1988, 1994) replicated Greenberg's study, and characterized the co-occurrence classes in terms of the combination of place of articulation and the major manner feature [sonorant]. Padgett (1995) proposed that the [−sonorant] segments are further subdivided into sections by the feature [continuant], but only for the coronal segments.

The traditional approach to the Arabic co-occurrence restrictions is to divide the consonants into natural classes, with co-occurrence constraints applying within these classes. The major co-occurrence classes discussed by Greenberg and McCarthy are presented in (2). In their analyses, consonants in any one of these classes are claimed to co-occur freely with consonants from any other class, and within any class consonants tend not to co-occur (except that the uvular fricatives /χ/ and /ʀ/ belong simultaneously to two different co-occurrence classes, the dorsals and the gutturals). Greenberg and McCarthy both note, however, that among the coronal obstruents, there are far more roots containing one fricative and one stop than roots containing two fricatives or two stops. They also note that co-occurrence between non-adjacent consonant pairs (in the first and third position of a triliteral root) is less restricted than co-occurrence between the other adjacent consonant pairs. The gradient nature of these generalizations suggests that an explanatory account of consonant co-

[2] As we discuss below, these exceptional cases of identity result from autosegmental spreading or copying, and these forms are actually derived from roots that contain only two consonants. At the level of representation where OCP-Place applies in the grammar of Arabic, these roots are not violations of the constraint.

occurrence in the Arabic roots must go beyond a categorical statement of co-occurrence classes.

(2)a. Labials = {b, f, m}

b. Coronal Obstruents = {θ, ð, t, d, tˤ, dˤ, s, z, sˤ, zˤ, ʃ}

c. Dorsals = {k, g, q, χ, ʁ}

d. Gutturals = {χ, ʁ, ħ, ʕ, h, ʔ}

e. Coronal Sonorants = {l, r, n}

In order to quantify over-representation and under-representation in the lexicon, we follow Pierrehumbert's (1993) use of O/E. The O/E measure uses the ratio of the observed number of occurring consonant pairs (O) to the number that would be expected if consonants combined at random (E). A value of O/E less than 1 indicates that there are fewer observed combinations than would be expected if consonants combined at random – that is, there may be a co-occurrence restriction affecting the consonants of the pair. If O/E is greater than 1, the number of pairs observed is greater than the number expected. O/E larger than 1 suggests that there is no co-occurrence restriction between consonants of the pair. Observed and expected rates of co-occurrence are derived from a phonological lexicon of 2,674 Arabic roots taken from a dictionary of standard Arabic (Cowan (1979), which is a later edition of the dictionary used by McCarthy in his studies). We assume that the lexical patterns of Standard Arabic reflect the lexical patterns of most Arabic dialects (Greenberg 1950). Due to their exceptional nature, roots with repeated second and third consonants were excluded (see section 2). This corpus is freely available from the first author as a computer file.

For an example of the O/E measure, consider Arabic triconsonantal roots of the form /d t C/ (where C is any consonant). Such roots are not found. Given the frequency of roots beginning in /d/ and the frequency of roots with /t/ in second position in our lexicon, 2.3 such roots are expected if consonants combine at random. In this case O/E = 0, the strongest degree of under-representation. There are 2 roots containing /d s C/, and 2.9 are expected, giving an O/E of 0.69 (under-representation). There are 4 roots with /d g C/ and 3.3 expected at random, giving an O/E of 1.21 (over-representation).

Table I shows the O/E for aggregated sets of consonants in adjacent (C1C2 or C2C3) and non-adjacent (C1C3) position in the root. Observed and expected counts were computed for individual consonant pairs, as in

TABLE I

Co-occurrence of consonant pairs in Arabic, aggregated by major class and distance. Major classes are shaded

		Labial	Cor obs		Dorsal	Guttural		Cor son
		b f m	t d tˤ dˤ	θ ð s z sˤ zˤ ʃ	k g q	χ ʁ	ħ ʕ h ʔ	l r n
Adjacent								
Labial	b f m	0.00	1.37	1.31	1.15	1.35	1.17	1.18
Cor obs	t d tˤ dˤ		0.14	0.52	0.80	1.43	1.25	1.23
	θ ð s z sˤ zˤ ʃ			0.04	1.16	1.41	1.26	1.21
Dorsal	k g q				0.02	0.07	1.04	1.48
Guttural	χ ʁ					0.00	0.07	1.39
	ħ ʕ h ʔ						0.06	1.26
Cor son	l r n							0.06
Nonadjacent								
Labial	b f m	0.30	1.08	1.02	1.26	1.25	1.28	1.11
Cor obs	t d tˤ dˤ		0.38	1.06	1.24	1.05	1.02	0.97
	θ ð s z sˤ zˤ ʃ			0.24	1.16	1.35	1.14	1.23
Dorsal	k g q				0.07	0.68	1.19	1.03
Guttural	χ ʁ					0.25	0.12	1.10
	ħ ʕ h ʔ						0.34	1.13
Cor son	l r n							0.67

the examples above, and then aggregated in three ways to produce this table. First, pairs in first and second position and in second and third position have been collapsed into a single set of adjacent consonant pairs. For example, for the root /d s w/, the pairs /d, s/ and /s, w/ are adjacent pairs. The pair /d, w/ is a non-adjacent pair. In addition, the order of consonants within the pairs has been ignored, so that the consonant pair /d, s/ is treated identically to the pair /s, d/. Finally, the data have been grouped into major classes to highlight the co-occurrence restrictions between the major classes and subclasses noted by Greenberg and McCarthy. The coronal obstruents are given separately in stop and fricative groups and the uvular fricatives are given separately from the dorsals and the gutturals to highlight the special behavior of these classes.

The posited co-occurrence restrictions within the major classes, shaded in gray, can be seen in Table I in the low O/E values for these combinations. For these classes, the number of observed consonant pairs is far below what is expected by chance, indicating that roots containing pairs within these classes are rare. Thus, the results of Greenberg (1950) and McCarthy (1994) are replicated in our lexicon. Note, though, that there is quantitative variation in co-occurrence levels across the different classes. As noted above, there are strong effects within the groups of coronal stops and coronal fricatives (O/E = 0.14 and 0.04, respectively for adjacent pairs), as compared to the effects between the coronal stops and fricatives (O/E = 0.52). Non-adjacent pairs in the major classes are also under-represented, but the strength of the co-occurrence restriction for non-adjacent pairs is much less than the strength of the co-occurrence restriction for adjacent pairs. In the case of combinations of coronal stop and fricative, non-adjacent combinations are not under-represented (O/E = 1.06).

Elmedlaoui (1995) discusses an additional sub-regularity in the consonant co-occurrence patterns. Greenberg (1950) originally pointed out that, among adjacent coronal stop-fricative combinations, the order fricative-stop (O/E = 0.63) is much more common than the order stop-fricative (O/E = 0.37). Elmedlaoui claims that other consonant pairs also show asymmetry. He proposes that OCP-Place violations with falling sonority are preferred to combinations with flat or rising sonority. The regularity based on sonority contour is also gradient, rather than categorical, providing additional evidence that a complete account of Arabic consonant co-occurrence must allow for gradient effects.

2. Previous Accounts of Arabic Morphology and OCP-Place

The first formalizations of the consonant co-occurrence restrictions in the Arabic verbal roots took advantage of the notation of autosegmental phonology. These analyses have attempted to account for two characteristics of the Arabic data. First, McCarthy (1986) proposed an account of the distribution of identical consonant pairs, e.g., *madad*, but **dadam*. Later work has tried to account for the under-representation of roots with consonant pairs in the major classes in any position (McCarthy 1994; Mester 1986; Padgett 1995).

Arabic verbal morphology can be represented by separating the vowels and consonants of the word form onto different autosegmental tiers (McCarthy 1979). Thus, a typical verb is represented as in (3), with each morphological tier contributing to the meaning of the word form.

(3)

```
vowel tier:            u     i
                       |     |
skeletal tier:      C  V  C  V  C
                    |     |     |
consonant tier:     k     t     b
```

McCarthy accounts for the co-occurrence restrictions in Arabic by application of the Obligatory Contour Principle, a constraint of Universal Grammar stated in (4), to the consonant root tier.

(4) OCP: Adjacent identical elements are prohibited

This rule is referred to as the 'total OCP' since it enforces the OCP for identical consonants. While Arabic does not allow roots where the first two consonants are identical, like */d d m/, it does permit the second pair of consonants to be identical, as in *madad*. The underlying root form of *madad* is taken to be /m d/ with only two consonants. Assuming association in Arabic proceeds from left to right, the proper surface form can be derived, with the resulting representation shown in (5).

(5)

```
C V C V C
|    \ /
m     d
```

Vowels are interspersed with the consonants giving surface forms where there appear to be two separate /d/s. McCarthy thus claims that triliteral roots with repeated second and third consonants are conceptually biliteral at the relevant level of abstraction, and do not violate the OCP.

Recent work in Optimality Theory has treated repeated final consonants as cases of reduplication rather than autosegmental spreading (Gafos 1998; Rose 2000). Rose (2000) claims that /m a d a d/ forms are still subject to the OCP but are allowed to violate it. This move is possible in Optimality Theory where constraints can be violated to satisfy higher ranked constraints.

According an even more recent OT treatment of Arabic morphophonology (Gafos 2001), the underlying representation of the verb stem in (5) is not the biliteral form /m d/, or the triliteral form /m d d/ but rather a form containing a final geminate, /m a d:/. The geminate 'separates' in the formation of (5), but appears intact elsewhere in the paradigm. The impossibility of forms containing an initial geminate (e.g., /m: a d/ as well as output candidates such as /m a m a d/) follows from the lack of prefixation in Arabic together with paradigm uniformity effects. In short, stems with doubled consonants are represented somewhere as geminates (i.e., as single elements on the melodic tier). The impossibility of absolutely word-initial geminates transfers to the rest of the paradigm.

The Gafos (2001) model shares with the original McCarthy treatment the property that forms with doubled consonants are only possible if the doubled consonant corresponds to a single element elsewhere (in the McCarthy model, elsewhere is at the level of morpheme structure, whereas in the Gafos model, elsewhere is elsewhere in the paradigm). This provides a way for the collapsed form of the doubled consonant to license an unpacked form. This morphophonological phenomenon, which pertains only to cases of total identity, is obviously categorical. In the descriptive terms of Figure 1, it would be represented as a step function located at the extreme edge of the diagram. Our formal toolkit indeed provides this option. Totally identical consonants are thus subject to two different constraints. Because total identity is the maximum value of similarity, totally identical consonants in stem forms such as (McCarthyite) /m d d/, /m m d/, /d d d/ or (Gafosian) /m d a d/, /m a m d/, or /d a d:/ are all ruled out by the gradient-similarity constraint we develop here. A separate and additional constraint categorically rules IN a doubled consonant corresponding to a geminate elsewhere in the paradigm and/or in underlying representation. In either the autosegmental or the OT account, this categorical permission overrides the gradient preclusion.

Issues concerning the mechanism for achieving this override are orthogonal to the present work, and thus we continue our discussion using the representational apparatus of autosegmental phonology. There is ample evidence for the psychological reality of the projection of consonant roots by abstracting over word forms. This data comes from speech production tasks (Berg and Abd-El-Jawad 1996), the performance of an aphasic patient (Prunet et al. 2000), and metalinguistic tasks (McCarthy 1986; Berent and Shimron 1997) that show the consonant root to be a level of representation in phonological processing by Arabic speakers. We discuss this data further in section 4.

To account for co-occurrence restrictions for non-identical consonants within the major classes, McCarthy (1988) applied the OCP to individual place feature tiers to rule out roots containing homorganic consonants in any position. This constraint is referred to as OCP-Place ('Adjacent identical place features are prohibited'). For example, a hypothetical root like */f t b/ is represented as in (6), with irrelevant feature tiers excluded for clarity. Adjacent identical features on the labial tier violate OCP-Place, marking the structure as ill-formed.

(6)
```
labial tier:    [lab]        [lab]
                  |            |
skeletal tier:   *C      C     C
                         |
coronal tier:          [cor]
```

McCarthy uses [labial], [coronal], [dorsal], and [pharyngeal] place features to divide the consonants into the major co-occurrence classes. The split of the coronals into two major classes is explained by special reference to the feature [sonorant] in the OCP-Place constraint for coronals. The uvulars {q, χ, ʁ} are assigned both [dorsal] and [pharyngeal] place, to account for the dual patterning of /χ/ and /ʁ/ with the dorsal and guttural sections. McCarthy (1994) argues that [pharyngeal] consonants are also split by manner, as /q/ does not have a co-occurrence restriction with the other pharyngeals and laryngeals. Finally, the split among the coronal obstruents into stops and fricatives could also be formalized by reference to the feature [continuant], following the analysis of Padgett (1995). Recent accounts in Optimality Theory have formalized the OCP using conjunction of markedness constraints such as *Place(coronal)^2 to avoid two specifications of [coronal] (Alderete 1997; MacEachern 1999). Note, however, that these proposals treat all differences in degree of co-occurrence identically. All violations are ungrammatical, regardless of which repeated feature is involved. As McCarthy admits, the split between the coronal obstruents and coronal sonorants is much stronger than the split between the coronal stops and fricatives, but they are accounted for with the same formal device.

Pierrehumbert (1993) pointed out a number of formal and empirical difficulties with the categorical OCP analyses that emerge when the quantitative co-occurrence data are examined in detail. First note that identical consonants are homorganic, and thus the OCP-Place constraint subsumes the total OCP for adjacent consonant pairs. However, McCarthy (1988, 1994) maintains the distinction between the total OCP and OCP-Place because the total OCP is stronger than OCP-Place. Adjacent identical consonants are prohibited. Roots with homorganic consonants do occur but they are highly under-represented. The difference between these two constraints is one of degree. McCarthy (1994) notes the distinction, but provides no formal account of the difference. Though OT allows constraints to be violated, the conjoined markedness constraints in OT specify a categorical input-to-output mapping. That is, the net result of adjudicating constraint violations is the same for all input forms that share the relevant particulars. In the analysis of Arabic, for example, a single consonant pair would be either acceptable or unacceptable in all word forms (Berkley 1994b; Pierrehumbert 1999; Plenat 1996). Under versions of OT that permit stochastic constraint ranking, such as Boersma and Hayes (2001), probabilistic variation in the outcomes would result in probabilistic variation for all words, not in differential probabilities in the lexicon. In addition, the cumulative interaction of similarity in all dimensions is problematic to formalize in OT. As discussed in Kirchner (1997), the OT architecture handles cumulative effects only on single dimensions, not on multiple dimensions. This problem is discussed in greater detail in Pierrehumbert (1999), as well as below in section 5.

Differences in the degree of under-representation, such as the combination of coronal stops and fricatives and the distinction between adjacent and non-adjacent pairs, are given no account in a categorical OCP analysis. In each case, an analysis using categorical co-occurrence restrictions must decide whether the degree of restriction is strong enough to warrant inclusion in the set of co-occurrence classes. These unexplained differences in degree of co-occurrence motivate our proposal for a quantitative similarity-based account.

3. OCP-Place as Similarity Avoidance

There is a clear sense in which the OCP effect is cumulative, with the total OCP as the strongest case of the more general OCP-Place constraint. This insight is present in Greenberg's (1950) discussion, but is absent from the categorical OCP account. Non-place features do have some role to play in the categorical OCP account, as both the total OCP and the division of the coronals into two or three co-occurrence classes by manner features involve reference to non-place features. Pierrehumbert (1993), extending an observation of Lightner (1973), proposed that any non-place feature is potentially relevant to the strength of the OCP-Place constraint. In other words, consonant co-occurrence can be accounted for through a single gradient constraint. She proposed that homorganic consonants are avoided in the Arabic roots as a function of their similarity. Identical consonants are maximally similar and have the strongest co-occurrence constraint. Consonants that differ in many features but are still homorganic are subject to weak co-occurrence restrictions. The influence of similarity on consonant co-occurrence is also affected by distance, as the constraint is weaker for non-adjacent consonants. In the remainder of this section, we present evidence supporting the similarity account of consonant co-occurrence in Arabic. We also present an explicit similarity metric that can capture many of the gradient co-occurrence patterns.

3.1. *Similarity Effects within the Major Classes (Sub-Classification)*

Some descriptive inadequacies of the categorical account of OCP-Place were originally pointed out in Pierrehumbert (1993). For example, she showed that the autosegmental model makes incorrect predictions on the effects of distance on the total OCP and OCP-Place. It has already been established that OCP-Place is stronger for adjacent consonant pairs than non-adjacent pairs. There is also an analogous effect of distance on the total OCP. However, in the autosegmental OCP account, the total OCP can only apply to adjacent consonants, as all feature tiers are relevant in establishing that two consonants are identical. For non-adjacent consonants, the intervening consonant will always have some features that share tiers with the surrounding segments, so there will always be a blocker between non-adjacent identical consonants on some tier. Therefore the total OCP cannot apply to non-adjacent segments in the autosegmental OCP account. Pierrehumbert (1993) showed that identical non-adjacent consonant pairs are more restricted (O/E = 0.14) than non-identical non-adjacent homorganic pairs (O/E = 0.62), contrary to the prediction of the autosegmental OCP.

Another difficulty for the categorical OCP account is the fact that the co-occurrence of coronal consonants depends on their manner of articulation. Greenberg originally pointed out that the coronal obstruents actually break into two classes, the coronal stops and coronal fricatives, as seen above. There are several other cases of sub-classification within the coronals. For example, the emphatic coronals /tˤ, dˤ, sˤ, zˤ/ have a stronger co-occurrence restriction with each other (O/E = 0 for adjacent pairs) than they do with the other coronal obstruents (O/E = 0.35). Among the coronal sonorants, /l/ and /r/ form a subclass, as they have stronger co-occurrence restrictions with each other (O/E = 0) than they do with /n/ (O/E = 0.15). Finally, voicing influences the co-occurrence of coronal obstruents. Pairs of coronal obstruents with the same voicing specification are found less frequently (O/E = 0.21) than obstruents with different voicing specifications (O/E = 0.36).

Outside of the coronal place class, there are also gradient effects of manner features on co-occurrence. Like the coronals, the [pharyngeal] group is split by manner. However, the split is not categorical. The dorsal stop /q/, which is specified for [dorsal] and [pharyngeal] place, co-occurs with the true gutturals /ħ, ʕ, h, ʔ/ (O/E = 0.85) less frequently than the non-pharyngeal dorsal stops /k, g/ do (O/E = 1.17). This suggests that there is a co-occurrence constraint between /q/ and the gutturals due to [pharyngeal] place. The labial class also shows some evidence of sub-classification by manner. While there are no observed pairs of labial consonants in adjacent position, there are 17 such pairs in non-adjacent position. Of those, 16 involve /m/ with a labial obstruent (/b/ or /f/). There is only one root with two non-adjacent labial obstruents, suggesting that manner does indeed have an effect on co-occurrence for labials as well. The effect of manner for labials is more difficult to observe than the effect of manner for coronals, as the co-occurrence constraint is much stronger within the labial class than within the coronal class.[3]

All of these cases of sub-classification are unexplained in a categorical OCP account. Further, it is important to note that all of these sub-classifications are not just partitions of the data into smaller groups of consonants that categorically cannot co-occur. For example, if the coronal obstruents are partitioned into two groups by use of the feature [continuant], many exceptions to a categorical OCP-Place constraint that does not employ [continuant] are ruled out. At first, this may seem to be an improvement in the account, as the large number of co-occurrences of coronal stops and fricatives would no longer be subject to the categorical OCP. However, categorically splitting the coronals by [continuant] introduces another kind of unaccounted for exception. The coronal stops and fricatives co-occur far less frequently than they should, so their systematic under-representation is no longer explained by this modified OCP-Place constraint.

Because the OCP-Place constraint in Arabic is gradient, dividing the Arabic consonants categorically into groups that can or cannot co-occur results in a trade-off between reducing the number of exceptions to OCP-Place and accounting for the under-representation of combinations that are not subject to the constraint. This trade-off is summarized in Table II. Table II shows the number of exceptions across the Arabic roots for OCP-Place constraints that are sensitive to different combinations of place and manner features. Exceptions are roots that occur in the lexicon despite containing consonant pairs that are subject to the categorical OCP. Also given in Table II is the amount of unexplained under-representation of homorganic consonant pairs outside of the co-occurrence classes that are defined by those same features. Unexplained under-representation is an estimate of the number of pairs of unrestricted consonant combinations that should occur in the lexicon, but do not. The number of expected combinations for unrestricted consonants is estimated using the O/E of 1.22 for non-homorganic pairs, such as in combinations of the labials and coronals. The

[3] All of the non-adjacent labial forms involve the pattern *b/f C m*. Elmedlaoui (1995) discusses the likely historical origin of these forms as the assimilation of an ancient /m/ suffix into the lexical roots. We hypothesize that the assimilation could take place precisely because the labial consonants were non-adjacent and maximally dissimilar, thus minimally violating similarity avoidance.

TABLE II

Trade-off between ruling out exceptional co-occurrence and explaining statistical under-representation in adjacent consonant pairs for different autosegmental models of feature co-occurrence. Exceptions are consonant pairs that co-occur despite sharing the features that define a co-occurrence class. Unexplained under-representations are missing consonant pairs that would be expected to co-occur given the rate of co-occurrence of unrestricted pairs (O/E = 1.22)

Definition of Classes	Exceptions	Unexplained under-representation
Place only	816	–
Place & [son]	123	160.8
Place & [son] & [cont]	36	312.7
Enumerated pairs	–	430.5

most general OCP-Place constraint, based only on place of articulation, leads to a large number of exceptions. As the natural classes for the OCP-Place constraint include more manner features and become more specific, smaller groups of consonant pairs are ruled out, and fewer exceptions are found. But by ruling out co-occurrence for very specific groups of consonants, the amount of unexplained under-representation in consonant pair co-occurrence increases. For example, including Place & [son] & [cont] in the OCP does not explain the under-representation of forms with a coronal stop-fricative pair like /tasaba/ (O/E = 0.52). The last case in the table, called ‘enumerated pairs’, does not base co-occurrence on a coherently defined natural class. Instead, the excluded consonant pairs are only those pairs that do not co-occur in adjacent position. While enumerating the individual impossible pairs leads to no exceptions to OCP-Place, the amount of unaccounted for under-representation is large. In other words, there are significant phonotactic generalizations that are missed completely.

Note also that Table II only presents data for adjacent consonant pairs. The problem of exceptions and unexplained under-representation in the autosegmental OCP becomes worse if non-adjacent pairs are included as well. The sonority-based asymmetry discussed by Elmedlaoui (1995) is an additional regularity where the differences in co-occurrence are not categorical. We conclude that any account that divides the consonants into categorical co-occurrence classes is unable to provide both a low rate of occurrence of exceptions and a low rate of unexplained under-representation. We feel that both of these dimensions of co-occurrence are important to consider, as it is undesirable to have an account with a large number of exceptions or an account that fails to capture the observable generalizations.

3.2. *Similarity Effects Between the Major Classes (Cross-Classification)*

In addition to sub-classification within the previously reported major classes, there are statistical regularities in co-occurrence outside of the major classes. McCarthy (1994) divides the Arabic consonants into four places of articulation: [labial], [coronal], and [dorsal] and [pharyngeal]. We saw above that the dual patterning of /χ, ʁ/ with the dorsals and gutturals can be explained if they are [dorsal] and [pharyngeal]. The emphatic coronals /tˤ, dˤ, sˤ, zˤ/ are another class of consonants that have two places of articulation. McCarthy characterizes these segments as uvularized, and assigns them secondary [pharyngeal] place with a redundant [dorsal] place specification. However, McCarthy does not apply the OCP to secondary place features. Kenstowicz (1994) observes that the emphatics would be better characterized with a secondary [dorsal] articulation that is active in the OCP (see also Bachra 2000). The emphatics have a fairly strong co-occurrence restriction with the velar stops /k, g/ (O/E = 0.13). There is no evidence of a co-occurrence restriction between the emphatics and the uvulars /q, χ, ʁ/ (O/E = 1.56) or the other gutturals /ħ, ʕ, h, ʔ/ (O/E = 1.23).

Despite the fact that the coronal obstruents and coronal sonorants have been divided into separate equivalence classes for OCP-Place, there is some evidence that their co-occurrence is restricted. In the previous section, we demonstrated a small effect of voicing on the co-occurrence of coronal obstruents. We have found a similar effect of voicing on the co-occurrence of coronal obstruents and coronal sonorants. Sonorants are phonetically voiced, though sonorant voicing is phonologically non-contrastive. Effects of voicing on co-occurrence between coronal obstruents and sonorants suggests that sonorant voicing is relevant to OCP-Place, despite the fact that it is phonologically redundant. The O/E for coronal sonorants and voiced coronal obstruents is 1.15 (239 actual and 207 expected). The O/E for coronal sonorants with voiceless obstruents is 1.31 (245 actual and 187 expected). The effect of redundant voicing is subtle and it is marginally statistically significant. To test statistical significance, coronal obstruents that contrast only by voicing can be paired with coronal sonorants, and comparisons made between pairs. For example, we can compare O/E for /t, n/ with /d, n/, /s, l/ with /z, l/, and so on. There are $n = 30$ pairs. The difference in O/E between voiced and voiceless pairs approaches significance using the Wilcoxon Signed-Rank test (Devore 1987; Signed-rank sum, $S_+ = 305$, $p = 0.07$). While the lack of strict statistical significance for this pattern might appear to be problematic, similarity avoidance predicts that very small differences in similarity will result in very small differences in rates of co-occurrence. Since statistical significance is based on an estimate of possible variation in a random sample, small differences in co-occurrence patterns will return weaker levels of statistical significance. So this small perturbation of the co-occurrence pattern can be accounted for if the effect of voicing on the similarity of obstruents and sonorants is predicted to be small as well.

The final pattern of co-occurrence outside of the major classes that we have discovered concerns the glides. First, there is some evidence for a co-occurrence restriction between the glide /w/ and the labials. This restriction appears to be sensitive to manner. The co-occurrence of /w/ and /m/ (O/E = 0.66) is more restricted than the co-occurrence of /w/ and /b, f/ (O/E = 0.93). In both cases, however, the co-occurrence is less than the co-occurrence of labials with other places of articulation (O/E = 1.22). For the purposes of OCP-Place, /w/ is a member of the labial class. There is also evidence for a co-occurrence restriction between /j/ and the velars /k, g/ (O/E = 0.62). The co-occurrence of /j/ with the other [dorsal] consonants (O/E = 1.09) is also a little lower than the co-occurrence of unrestricted consonants (O/E = 1.22). Though the differences are slight, this pattern appears to have some additional sub-regularity. For example, the co-occurrence of /j/ with the /k/ is much more frequent (O/E = 0.90) than the co-occurrence of /j/ with the /g/ (O/E = 0.37), which in many Arabic dialects is realized as a voiced palatal affricate.

The co-occurrence restrictions outside of the major classes that are demonstrated in this section are weaker than the effects found within the major classes, but they are consistent with the similarity account. A categorical OCP model has no appropriate tool for capturing these restrictions, however. There is no principled way for it to predict small differences in degree of co-occurrence. In a categorical OCP account there is a choice of getting no effect, or a uniformly strong effect. Using similarity, we can get a systematic range of intermediate effects and account for gradient patterns of cross-classification among the Arabic consonants.

3.3. *Contrast, Feature Specification, and Similarity*

The autosegmental OCP analysis used two separate constraints to account for the difference in degree of co-occurrence between identical consonant pairs and homorganic but non-identical consonant pairs. In order to account for the additional cases of sub-classification, the categorical OCP analysis requires an explosion of individual sub-constraints. In theory, each of these sub-constraints could be categorical or a tendency. In the similarity account, all of the co-occurrence restrictions can be captured by a single constraint, similarity avoidance, where the degree of co-occurrence restriction depends on the similarity between homorganic consonants. The strength of the constraint is predicted by the similarity of the consonants that are involved, and not on an *ad hoc* basis. For all cases of OCP co-occurrence restrictions that have been studied so far, the similarity prediction is supported, as discussed in section 4.2.

Conceptually, a similarity metric for phonological segments is not a difficult problem, as phonological features provide a ready basis upon which to compute similarity (e.g., van den Broeke and Goldstein 1980). However, there are two aspects of the co-occurrence restrictions of Arabic that are challenging to an explicit similarity account. In the autosegmental OCP account, manner features were applied to some places of articulation but not others. Manner features were used selectively because manner is more relevant to the co-occurrence of coronals and pharyngeals than it is to the co-occurrence of labials and dorsals, as can be seen in Table II. Pierrehumbert (1993) observed that the larger place classes are the ones that are clearly divided into manner sub-classes, while the smaller place classes provide much less evidence for an effect of manner. Thus, similarity must be a function of the size of the inventory at each place of articulation. Following Pierrehumbert (1993), we propose that similarity is dependent upon contrast. Larger classes, with more contrasts, are the ones that divide into smaller sub-classes.

We propose to compute similarity over the natural classes of a segment inventory, rather than over the features directly. In the natural classes similarity metric, consonants with shared features are still similar, as the set of natural classes is derived from the set of features and segments using set theory (Broe 1993). But natural classes also reflect contrast within a segment inventory, so a similarity metric based on natural classes is sensitive to contrastiveness. Features that are not contrastive do not define unique natural classes, and so will not contribute to similarity. Further, features that are partially contrastive, like voicing across obstruents and sonorants, will in some cases define contrastive natural classes (e.g., combinations of [voice] and [continuant]) and in other cases will not define natural classes (e.g., combinations of [voice] and [+sonorant]). Thus, partially contrastive features will have some influence on similarity, but not as much influence as features that are fully contrastive. The natural classes similarity metric provides a simple solution to the more general problem of determining similarity for non-orthogonal category systems. For further discussion of the representation of a segment inventory using natural classes, see Broe's (1993) theory of structured specification.[4]

We compute similarity by comparing the number of shared and unshared natural classes of two consonants, using the equation in (7). This equation is a direct extension of the Pierrehumbert (1993) feature similarity metric to the case of natural classes.

(7) $$\text{Similarity} = \frac{\text{Shared natural classes}}{\text{Shared natural classes + Non-shared natural classes}}$$

Identical consonants have similarity 1, as they participate in exactly the same natural classes. Highly dissimilar consonants share very few natural classes (because they share very few features) and have very low similarity. Since OCP effects in Arabic only apply to consonants that share major place of articulation features, we stipulate that the natural classes used in the similarity computations are only those natural classes containing a place of articulation feature. Thus, nonhomorganic consonant pairs will have similarity 0, as they share no natural classes containing a place feature.[5]

[4] Pierrehumbert (1993) accounted for the effect of inventory size on similarity using *contrastive underspecification* (Steriade 1987). Features that were phonetically realized, but not contrastive within a place of articulation, were phonologically unspecified and did not influence similarity. For example, the large coronal class needs more contrastive features, and so consonant pair similarity in the coronal class is lower due to a larger number of unshared features between most coronal consonants. Thus, Pierrehumbert achieved the influence of contrast on similarity by manipulating how features are specified. Pierrehumbert's solution parallels work in psychology on similarity and categorization that also treated non-contrastive features as though they were non-existent for the similarity computation (e.g., Tversky 1977). We present an alternative to the underspecification approach to similarity and contrastiveness that is superior on both empirical and theoretical grounds. We believe that all distinctive features influence similarity to some degree, as in the case of the influence of redundant voicing on the co-occurrence of the coronal obstruents and sonorants that is discussed above. More generally, eliminating contrastive underspecification is desirable as there is little empirical support for it elsewhere in phonological theory. Contrastive underspecification also suffers from serious formal shortcomings as outlined in Broe (1993) and Steriade (1995). We believe a similarity metric using natural classes provides an improvement over the use of contrastive underspecification, as the natural classes representation has none of the formal disadvantages of underspecification. The influence of contrast on similarity is grounded in the similarity metric itself, and not the choice of feature specifications for the inventory.

[5] The seemingly arbitrary restriction of similarity avoidance to place-based classes in Arabic is an example of a more general pattern in phonetically-motivated phonology. It is impossible for any language to satisfy every phonetically-motivated constraint. The functional pressures that lead to similarity avoidance in Arabic are released along the place dimension only, and not in manner or voicing (though Bachra (2000) clams there is also a weak manner-based restriction in the Arabic roots). This issue is discussed further in section 4.

We illustrate the metric with sample similarity values for labial consonants, using the distinctive features given in (8), adapted from Kenstowicz (1994). The natural classes given are only those whose definition includes the labial place feature. The pair /f, m/ share 2 such natural classes, namely {b, f, m, w} (the labials) and {b, f, m} (labial consonants). They have 7 non-shared classes: {b, f} (obstruents), {f, w} (continuants), {f} (voiceless continuants), {b, m, w} (voiced), {b, m} (voiced stops), {m, w} (voiced sonorants), and {m} (nasals). The similarity of /f, m/ is 2/9 = 0.22 by equation (7). The pair /b, f/ share 3 classes, {b, f, m, w}, {b, f, m} and {b, f}. They have 5 non-shared classes: {f, w}, {b, m, w}, {b, m}, {b, w}, {b}. The similarity of /b, f/ is 3/8 = 0.38.

We claim that similarity for coronal pairs like /s, n/ is less than the similarity of /f, m/ due to the larger space of contrasts in the coronals. As a result, the co-occurrence restriction for /f, m/ is stronger than the co-occurrence restriction for /s, n/. To see that this does indeed follow from structured specification, consider a hypothetical labial inventory with more contrasts. For example, suppose the Arabic inventory were to contain two additional labials, /p/ and /v/. The addition of these two segments adds several natural classes to the labials based on new contrasts in voicing and continuancy, even though there is no need to add distinctive features to describe the larger inventory. In the expanded inventory, the pair /f, m/ still share 2 classes: {p, b, f, v, m, w} (the labials) and {p, b, f, v, m} (the labial consonants). They now have 11 non-shared classes, so the similarity in the expanded inventory is 2/13 = 0.15. The four additional non-shared classes are: {f, v} (the new class of continuant obstruents without regard to voicing), {p, f} (the new class of voiceless obstruents without regard to continuancy), {p, b, m} (the new class of stops without regard to voicing), {b, v, m} (the new class of voiced consonants without regard to continuancy). Using a similarity metric based on natural classes, the difference in the strength of OCP effects in the major classes can be explained without resorting to underspecification of phonetically relevant features or parametric conditions on the co-occurrence constraints. The larger number of natural classes among the coronals, and to some extent the dorsals and pharyngeals, decreases the similarity between consonant pairs in those classes that share few features. In a small class, like the Arabic labials, these non-contrastive features do not contribute to dissimilarity. Crucially, the set of classes that count and do not count toward similarity is determined uniquely by algorithm from the set of segments and features in the inventory. A more extensive discussion of the natural classes similarity metric can be found in Frisch (1996), using the formalism of structured specification developed in Broe (1993).

The complete feature matrix used to generate natural classes and compute similarity for the Arabic consonants is given in (8). Similarity values for each consonant pair are given in Table III. In general, the use of contrastive underspecification is avoided in these feature assignments. However, some feature assignments have been left unspecified where the features are inapplicable or irrelevant, a pattern dubbed *trivial underspecification* by Steriade (1995). For example [strident] is specified within the coronal fricatives. It is inapplicable to non-fricatives, because its phonetic definition (involving direction of a turbulent air stream against the teeth) presupposes the existence of a turbulent air stream. This presupposition is only met for fricatives. The feature [−strident] could also be specified for all non-coronal fricatives, but this would have no effect on the similarity metric so this feature specification (or lack thereof) has no relevance in this study. Of course, there are a number of alternative ways that features and their specifications could be chosen for the Arabic inventory. The features in (8), taken from Kenstowicz (1994), provide a sufficient basis to generate the many natural classes that are relevant to OCP-Place. We use binary features (such as [+voice] and [−voice]) as a familiar way to represent a bivalent contrast. However, the set theoretic basis of the natural classes used in structured specification implies that the metric would produce the same results if all features were monovalent (e.g., [+voice] and [+unvoiced]). By using natural classes, the similarity metric is stable under relabeling of features (see Frisch et al. 1997). Specification of major place features follows McCarthy (1994) except in one case: The emphatic coronal consonants are specified as [dorsal] rather than [pharyngeal]. This particular assignment of place features accounts for the co-occurrence restriction between the emphatics and the velars as argued by Kenstowicz (1994) and Bachra (2000).

(8)	b	f	m	t	d	tˤ	dˤ	θ	ð	s	z	sˤ	zˤ	ʃ	k	g	q	χ	ʁ	ħ	ʕ	h	ʔ	l	r	n	w	j
cons	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−	−	+	+	+	−	−
son	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+
cont	−	+	−	−	−	−	−	+	+	+	+	+	+	+	−	−	−	+	+	+	+	+	−	+	+	−	+	+
acute				+	+	−	−	−	−	+	+	−	−	−														
strident								−	−	+	+	+	+	+														
nasal	−		+	−	−	−	−								−	−	−									+		
lateral																								+	−			
labial	+	+	+																								+	
coronal				+	+	+	+	+	+	+	+	+	+	+										+	+	+		
anterior				+	+	+	+	+	+	+	+	+	+	−										+	+	+		
dorsal						+	+					+	+		+	+	+	+	+									+
back						+	+					+	+		−	−	+	+	+									−
phar																	+	+	+	+	+	+	+					
radical																				+	+							
voice	+	−	+	−	+	−	+	−	+	−	+	−	+	−	−	+	−	−	+	−	+			+	+	+	+	+
spglot																						+						
conglot																							+					

There are two other assignments of features that are worthy of mention. First, the [acute] feature is specified for all coronal obstruents. This perceptual feature is intended to encode a contrast in the frequency spread of the stop burst, aspiration, and/or frication of coronal obstruents. This contrast serves primarily to differentiate the plain coronals from their emphatic counterparts for stops and fricatives. Without it, or some equivalent feature, there is no natural class of non-emphatic coronals distinct from the emphatic coronals. Note that it is not sufficient to consider the non-emphatic coronals to be individuated by their lack of the secondary [dorsal] articulation. This would use an unspecified feature as equivalent to a feature specification such as [−dorsal]. The specification of contrasts by the covert use of the absence of a feature leads to a number of formal difficulties that have been discussed by Stanley (1967) and Broe (1993).

The second feature to note is the feature [lateral]. This feature is only specified for /l/ and /r/. In (8), [lateral] is used as a cover feature for whatever combination of properties differentiates /l/ from /r/. From the perspective of a natural classes similarity metric, the exact feature or features used and their labels is not crucial. In this case, a single feature serves to define the appropriate natural classes, and it is the natural classes rather than the features that are counted in the similarity metric.

3.4. *Quantitative Comparison of Accounts*

We have already shown qualitatively that OCP-Place effects in Arabic include much sub-regularity that suggests a gradient, similarity-based account. In this section, we develop such an account explicitly by demon-

TABLE III

Similarity for Arabic consonants using the natural classes metric

	b	f	m	t	d	tˤ	dˤ	θ	ð	s	z	sˤ	zˤ	ʃ	k	g	q	χ	ʁ	ħ	ʕ	h	ʔ	l	r	n	w	j
b	1	0.38	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.22	0
f	0.38	1	0.22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0
m	0.5	0.22	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.38	0
t	0	0	0	1	0.42	0.26	0.17	0.21	0.14	0.32	0.19	0.12	0.08	0.12	0	0	0	0	0	0	0	0	0	0.1	0.1	0.18	0	0
d	0	0	0	0.42	1	0.16	0.3	0.13	0.21	0.17	0.32	0.08	0.13	0.07	0	0	0	0	0	0	0	0	0	0.15	0.15	0.31	0	0
tˤ	0	0	0	0.26	0.16	1	0.4	0.24	0.14	0.14	0.09	0.41	0.21	0.13	0.21	0.11	0.36	0.12	0.07	0	0	0	0	0.05	0.05	0.09	0	0.03
dˤ	0	0	0	0.17	0.3	0.4	1	0.13	0.23	0.09	0.14	0.2	0.42	0.07	0.11	0.24	0.17	0.07	0.14	0	0	0	0	0.09	0.09	0.16	0	0.06
θ	0	0	0	0.21	0.13	0.24	0.13	1	0.45	0.4	0.24	0.45	0.24	0.37	0	0	0	0	0	0	0	0	0	0.15	0.15	0.07	0	0
ð	0	0	0	0.14	0.21	0.14	0.23	0.45	1	0.25	0.44	0.24	0.44	0.21	0	0	0	0	0	0	0	0	0	0.26	0.26	0.13	0	0
s	0	0	0	0.32	0.17	0.14	0.09	0.4	0.25	1	0.44	0.35	0.2	0.3	0	0	0	0	0	0	0	0	0	0.16	0.16	0.08	0	0
z	0	0	0	0.19	0.32	0.09	0.14	0.24	0.44	0.44	1	0.2	0.35	0.17	0	0	0	0	0	0	0	0	0	0.27	0.27	0.13	0	0
sˤ	0	0	0	0.12	0.08	0.41	0.2	0.45	0.24	0.35	0.2	1	0.42	0.33	0.11	0.06	0.17	0.15	0.09	0	0	0	0	0.09	0.09	0.04	0	0.04
zˤ	0	0	0	0.08	0.13	0.21	0.42	0.24	0.44	0.2	0.35	0.42	1	0.17	0.07	0.13	0.09	0.1	0.21	0	0	0	0	0.14	0.14	0.07	0	0.09
ʃ	0	0	0	0.12	0.07	0.13	0.07	0.37	0.21	0.3	0.17	0.33	0.17	1	0	0	0	0	0	0	0	0	0	0.09	0.09	0.05	0	0
k	0	0	0	0	0	0.21	0.11	0	0	0	0	0.11	0.07	0	1	0.38	0.32	0.12	0.07	0	0	0	0	0	0	0	0	0.12
g	0	0	0	0	0	0.11	0.24	0	0	0	0	0.06	0.13	0	0.38	1	0.15	0.07	0.15	0	0	0	0	0	0	0	0	0.24
q	0	0	0	0	0	0.36	0.17	0	0	0	0	0.17	0.09	0	0.32	0.15	1	0.32	0.15	0.08	0.04	0.04	0.09	0	0	0	0	0.04
χ	0	0	0	0	0	0.12	0.07	0	0	0	0	0.15	0.1	0	0.12	0.07	0.32	1	0.42	0.23	0.13	0.14	0.1	0	0	0	0	0.13
ʁ	0	0	0	0	0	0.07	0.14	0	0	0	0	0.09	0.21	0	0.07	0.15	0.15	0.42	1	0.12	0.17	0.14	0.09	0	0	0	0	0.27
ħ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.08	0.23	0.12	1	0.55	0.5	0.27	0	0	0	0	0
ʕ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0.13	0.17	0.55	1	0.56	0.3	0	0	0	0	0
h	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0.14	0.14	0.5	0.56	1	0.38	0	0	0	0	0
ʔ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.09	0.1	0.09	0.27	0.3	0.38	1	0	0	0	0	0
l	0	0	0	0.1	0.15	0.05	0.09	0.15	0.26	0.16	0.27	0.09	0.14	0.09	0	0	0	0	0	0	0	0	0	1	0.8	0.33	0	0
r	0	0	0	0.1	0.15	0.05	0.09	0.15	0.26	0.16	0.27	0.09	0.14	0.09	0	0	0	0	0	0	0	0	0	0.8	1	0.33	0	0
n	0	0	0	0.18	0.31	0.09	0.16	0.07	0.13	0.08	0.13	0.04	0.07	0.05	0	0	0	0	0	0	0	0	0	0.33	0.33	1	0	0
w	0.22	0.25	0.38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
j	0	0	0	0	0	0.03	0.06	0	0	0	0	0.04	0.09	0	0.12	0.24	0.04	0.13	0.27	0	0	0	0	0	0	0	0	1

TABLE IV

Co-occurrence of consonant pairs, aggregated by natural classes similarity, for adjacent and non-adjacent pairs

Similarity	Adjacent			Non-adjacent		
	Obs	Exp	O/E	Obs	Exp	O/E
0	4222	3456.8	1.22	1909	1715.8	1.11
0–0.1	484	459.6	1.05	252	247.2	1.02
0.1–0.2	378	453.8	0.83	226	231.3	0.98
0.2–0.3	167	282.7	0.59	102	130.7	0.78
0.3–0.4	91	281.3	0.32	139	154.2	0.90
0.4–0.5	3	92.5	0.03	10	29.9	0.25
0.5–0.6	2	31.7	0.06	9	19.1	0.47
0.8	0	53.2	0	11	22.6	0.49
1	1	236.5	0.01	16	113.0	0.14

strating that it provides a better quantitative fit to the patterns in the Arabic lexicon than a categorical account. We also consider an autosegmental account that uses a soft OCP-Place constraint that predicts the distribution of OCP-Place violations to be random. By contrast, the similarity-based account predicts that the number of OCP-Place violations decreases systematically as the similarity between the consonants involved increases. Thus, the similarity-based account displays the explanatory strength of a gradient constraint over a (merely) soft constraint.

Table IV shows the Observed counts, Expected co-occurrence, and O/E ratios for adjacent and non-adjacent consonant pairs as a function of natural classes similarity as given in Table III. Similarity of 0 for non-homorganic pairs and similarity of 1 for identical pairs are given separate entries. Similarity 0.8 for /l, r/ is also given a distinct entry. Other similarity levels in the range 0-0.6 are grouped arbitrarily into intervals of 0.1. The left side of Table IV shows the data for adjacent pairs, and the right side of Table IV shows the data for non-adjacent pairs. Clearly, there is a gradient effect of natural class similarity on co-occurrence, as the O/E ratios decrease as similarity increases. That the effect of similarity on co-occurrence weakens with distance is apparent when the O/E for adjacent consonant pairs is compared to the O/E for non-adjacent consonant pairs for each level of similarity.

However, the relationship between natural classes similarity and co-occurrence is not perfect. It should first be noted that, since the similarity metric is symmetric, the similarity account does not predict asymmetries in consonant co-occurrence (Pierrehumbert 1993, Elmedlaoui 1995). We discuss the significance of the asymmetries in section 5. A study of Table IV and the similarity matrix in Table III reveals several cases where the model could be improved. For example, for some pairs of coronal obstruents and coronal sonorants the computed similarity is too high. The similarity of 0.31 between /d/ and /n/ is particularly problematic. Its effects can be seen most clearly in the non-adjacent data, where the large number of /d/ and /n/ pairings bring the aggregate O/E out of line with the rest of the data. On the other hand, the similarity of /n/ to /l, r/ is too low. So the natural classes similarity metric is only partly successful in capturing the split of the coronals into obstruent and sonorant classes. The similarity of /w/ to the other labials is also too high, as their co-occurrence is only weakly restricted. Finally, /χ/ and /ʁ/ are not similar enough to the velars or to the other gutturals with which they have a strong co-occurrence restriction. It might be possible to resolve these inadequacies by weighting the features used in the similarity metric, as suggested by Bachra (2000). In particular, if the major manner features such as [sonorant] were weighted more than other features, all of the problematic similarity values would be improved. Such a weighting is perceptually and cognitively plausible. The major manner features are the most salient consonant features in speech (see Wright 1996 for a recent summary). Many psychological models of similarity employ weighting of features or dimensions of contrast based on salience (Medin and Shaeffer 1978; Nosofsky 1986). Weighting introduces additional parameters into the model, a move that can be justified in the psychological literature by experimental data demonstrating differential salience of features. Lacking such data, we have preferred to keep the number of free parameters in our model to a minimum.

To assess the extent to which the natural classes similarity model can capture the trends in the OCP-Place data, it was used to fit a simple decreasing function to the rate of occurrence in the lexicon of all consonant pairs. The model was fit separately to the adjacent and non-adjacent consonant pairs in the triliteral roots. There was one data point for each pair of consonants in first and second, second and third, or first and third position in the root. This is an ordered pair (x, y), where x is the similarity of the consonants in the pair and y is O/E for that pair. The model provides the best fitting prediction for O/E at each level of similarity, subject only to the restriction that the predicted O/E is monotonically decreasing (i.e., it does not increase at any point with an increase in similarity). This type of fit makes it possible to inspect the trend of the data without making specific assumptions about the exact form of the function involved. Goodness of fit

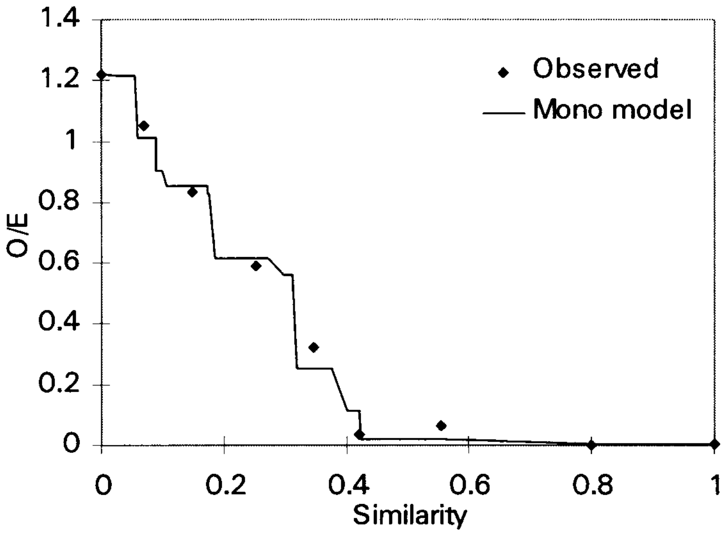

Figure 2. Aggregate O/E for adjacent pairs and a similarity model based on a decreasing function.

is evaluated by a residual sum of squares between the observed data and the model's predictions. The model was fit by first calculating the best O/E for similarity 0 pairs. Then the pairs with the smallest non-zero similarity were examined. If the best O/E for these pairs was higher than the similarity 0 pairs, they were grouped with similarity 0. If the best O/E for these pairs was lower, then they were separated into their own group. The fitting continued in the same manner as similarity of the pairs increases, with each new pair added to the previous or starting a new group depending on whether its O/E was higher or lower. Figure 2 shows the aggregated O/E for adjacent consonant pairs (as given in Table IV) along with the predicted O/E from the decreasing similarity function that provided the best fit.

In order to further demonstrate the empirical superiority of our account to previous analyses, we compare its ability to predict the occurrences of each type of consonant pair in each position in the verbal roots of Arabic against four other models of OCP-Place. Our model has two distinctive components: the frequency of co-occurrence is predicted by similarity, and similarity of consonant pairs is computed using natural classes. We refer to our model as the *natural classes model*.

An alternative model of the phonotactics of Arabic might accept the hypothesis that co-occurrence is a quantitative function of similarity, but reject the use of the natural classes similarity metric. An obvious alternative would be to compute similarity over features, rather than natural classes. A feature similarity model using the features in (8) is included to test the hypothesis that similarity is best computed over natural classes. This model directly tests the effect of contrast and redundancy in distinctive features in predicting co-occurrence in Arabic. We refer to this model as the *feature model*.

We created two other models of the data based on the autosegmental OCP account. The *categorical model* uses a categorical co-occurrence restriction on consonants in the major classes given in (2), whether adjacent or non-adjacent. This model predicts that there are no pairs of consonants from the major classes in adjacent or non-adjacent position. The categorical model represents a strict interpretation of the formal autosegmental account of the total OCP and OCP-Place effects. This model includes the special reference to [sonorant] for the coronal and pharyngeal consonants. Secondary articulations are not relevant to OCP-Place, following McCarthy (1994). Under this model, the total OCP is redundant given the OCP-Place constraint. Adjacent identical consonants are also homorganic, and thus are categorically ruled out by OCP-Place without the need for a total OCP. We note that this represents a literal interpretation of the formalism of the autosegmental OCP, including weaknesses that are already admitted by the authors in their own discussion of their analyses. However, we evaluate it because it provides an important baseline for the other OCP models. The fact that the categorical account was not intended to make quantitative predictions is seen below in the fit of the model. We also implemented a second autosegmental model that more faithfully reflects the discussion of OCP-Place in McCarthy (1988, 1994). This model uses a categorical co-occurrence restriction on identical consonants in adjacent position, and also a soft constraint against homorganic consonant pairs in adjacent and non-adjacent positions. The soft constraint is implemented as a constant predicted O/E value for all consonant pairs in the major classes given in (2). As in the categorical model, the feature [sonorant] sub-divides the coronal and pharyngeal classes, and secondary articulations are not subject to OCP-Place. We call this model the *soft model*.

Finally, we include for comparison a fifth model that does not take into account OCP-Place effects at all. We refer to this model as the *frequency model*, as consonant co-occurrence is predicted based only on the random combination of consonant pairs. In other words, it predicts O = E and the co-occurrence of consonant pairs that are OCP-Place violations or not OCP-Place violations is entirely random. The frequency model includes no provisions for the OCP-Place constraint, and thus serves to highlight the effect of implementing OCP-Place in the other models. In addition, this model makes explicit the contribution of expected consonant co-occurrence as a baseline measure of co-occurrence in the other models.

TABLE V

Model comparison. Residual sum of squares is shown separately for homorganic and major class data

Model	R^2	Residual SS	Hom. SS	Maj Class SS	# Param	Model Parameters
Frequency	0.57	14,476	8,697	7,101	0	O = E
Categorical	0.70	10,008	4,805	3,189	2	O/E = 0 for homorganic, O/E = 1.17 otherwise
Soft Model	0.73	8,918	3,716	2,100	3	O/E = 0 for adjacent ident, O/E = 0.38 for maj class, O/E = 1.17 otherwise
Feature Model	0.71	9,737	4,573	3,018	8	O/E = 1.20 to 0
Natural Classes	0.75	8,489	3,286	1,335	11	O/E = 1.22 to 0

All of the other models predict OCP-Place effects as a deviation from the expected co-occurrence given in the frequency model.

We compare all of the models based on their best fits to the data. For the two models that utilize similarity, the feature model and the natural classes model, we use the decreasing function that provides the best fit to the consonant pair data as described for Figure 2. We implemented the two autosegmental OCP models using constant predicted O/E values for each distinct sub-case of co-occurrence in the model. For the categorical model, predicted O/E for OCP-Place violations was 0, and a best fit constant value was used for non-violations. For the soft model, predicted O/E was 0 for total OCP violations, a best fit constant for OCP-Place violations, and a best fit constant for non-violations. The frequency model has no best fit parameters, predicted O = E for all consonant pairs.

Table V shows the models, their parameters, and the evaluations of goodness of fit. Overall fit of each model is given using the R^2 measure (a measure of the proportion of variation in the data that is accounted for) and the residual sum of squares (a measure of the amount of error in the model fit). In addition, we present two sub-cases of the residual sum of squares for each model. The first sub-case is the residual sum of squares over all homorganic consonant pairs, as defined by shared place features in the feature matrix in (8). The second sub-case is the residual sum of squares over just the consonant pairs within the major classes in (2).

Comparing R^2 values, the natural classes model provides the best overall fit to the data, accounting for 75% of the variation found in the data.

The monotonic decreasing similarity function used 11 different similarity levels to achieve the optimal fit, highlighting the gradience of the data. The frequency model shows that a significant percentage of the variation in the data is accounted for by frequency alone (57%), emphasizing the relevance of frequency as a base predictor of co-occurrence rate (Pierrehumbert 1994; Frisch 1996). Examining the residual sums of squares, note that much of the lack of fit for the frequency model comes from the homorganic pairs, which are the pairs that are subject to the OCP-Place constraint. In general, the models of OCP-Place perform equally well on the non-homorganic pairs, and the differences in fit are due to differences in fit for homorganic pairs.

The categorical model has the worst fit among the OCP-Place models. This model rules out all pairs in the major classes, whether adjacent or non-adjacent. The soft model performs much better, but because the categorical model and soft model cannot differentiate homorganic similar consonants from homorganic dissimilar consonants, they are unable to fit the homorganic consonant data as well as the natural classes model. Comparing the residual sum of squares over the major classes reveals that the natural classes model has superior performance over the core data set for the OCP-Place constraint. The difference in fit over the major classes is particularly striking as the performance of the autosegmental OCP models is based on parameters that are specifically fit to just the subset of the data identified by McCarthy (1994) as involving a phonological constraint. The natural classes model parameters come from an overall fit of all consonant pairs based on their similarity level as derived from our feature assignments.

In comparison to the other models, the feature similarity model was not very successful at capturing the sub-regularities within the major classes. The poor fit of the feature model to the major classes highlights the great advantage of natural classes similarity over feature similarity, demonstrating that contrastiveness does indeed play an important role in sub-dividing homorganic consonant groups by similarity. Insofar as the feature-based model of Pierrehumbert (1993) was successful, its success comes from its intuitively constructed system of contrastive underspecification for phonemes that simulated the contrastiveness effects that are provided rigorously and automatically by the natural classes model.

The major objection to the natural classes model just presented is that it involves an excessive number of parameters, many more parameters than the models it is competing with. These parameters arise because each step in the monotonically decreasing function provides a fresh choice of the step level. In general, one does expect the fit of a model to improve as more free parameters are provided. However, now that the pattern of

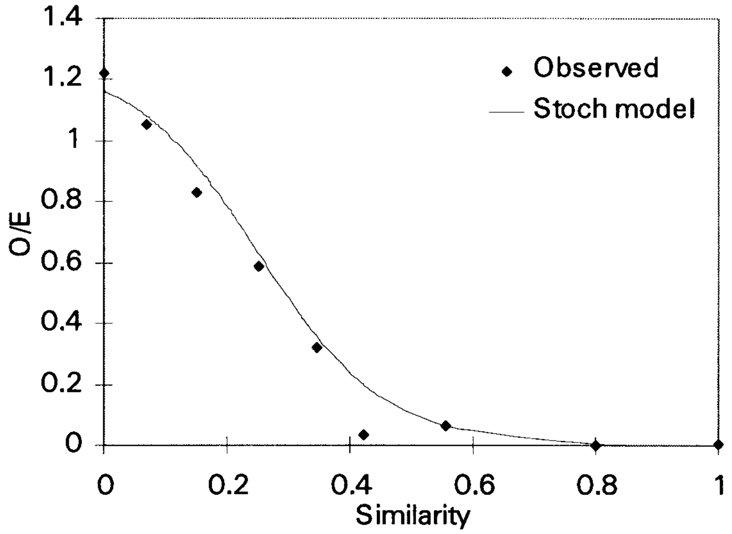

Figure 3. Aggregate O/E for adjacent pairs and a similarity model based on the stochastic constraint.

the effect has been clarified, we can return to the much lower parameter characterization introduced in Figure 1. This is the stochastic constraint model of Frisch, Broe, and Pierrehumbert (1997) using natural classes similarity. In this model, co-occurrence decreases with similarity but the relation between co-occurrence and similarity is constrained by a logistic function with only two free parameters, K and S. The stochastic constraint model achieved fits that were nearly as good as the decreasing similarity function that used many parameters. For the stochastic constraint model $R^2 = 0.744$, Residual SS = 8581, Homorg SS = 3455, and Maj Class SS = 1566. The best fit parameters were $K = -2.46$ and $S = 9.79$. Figure 3 shows this fit along with the aggregate data from Table IV.

In summary, the autosegmental OCP models only allow one or two levels of co-occurrence for homorganic consonant pairs, and so do not account for the full pattern of gradience. The natural classes similarity model describes the gradient co-occurrence pattern in a constrained way that accounts for many of the observed sub-regularities. Using logistic functions, this gradience can be summarized using only two free parameters. Not only does the similarity model capture gradience in an incisive way, it also successfully predicts differences in degree of co-occurrence from the intrinsic sensitivity of the natural classes similarity metric to contrastiveness.

4. Gradient Phonotactics in Phonology

The presence of robust quantitative generalizations in Arabic suggests that the phonotactic knowledge of Arabic speakers is more complex and detailed than is usually assumed in generative phonology. In this section, we support this hypothesis in three ways. First, evidence is presented to demonstrate that the gradient phonotactic constraints in the Arabic lexicon are psychologically real to Arabic speakers. Second, cross-linguistic evidence is examined that suggests gradient, similarity-based phonotactic constraints are common in the world's languages. Third, a theory of the phonetic origin of the similarity avoidance constraint in the Arabic lexicon is presented that suggests that OCP-Place patterns in the phonotactics are a reflection of processing constraints that shape the lexicon diachronically.

4.1. *The Psychological Reality of OCP-Place*

External evidence that a gradient OCP-Place constraint is an active component of the phonotactic knowledge of Arabic speakers comes from two sources. We review each in turn. Frisch and Zawaydeh (2001) conducted an experiment using nonce verb forms that contained a mixture of OCP-Place violations and non-violations. Thirty native speakers of Arabic living in Amman participated in the experiment. Frisch and Zawaydeh found that the participants judged novel verbs containing OCP-Place violations to be significantly less wordlike than novel verbs that contain no violations. In an analogous set of experiments, Berent and Shimron (1997) demonstrated the psychological reality of the OCP-Place constraint in Hebrew. Frisch and Zawaydeh also found some evidence that non-grammatical lexical factors influenced wordlikeness judgments, such as the availability of an analogy to occurring root forms and the frequency of occurrence of consonant pairs in the lexicon.

In a special set of stimuli that specifically controlled for these lexical factors, Frisch and Zawaydeh found that nonce verbs containing consonant pairs that are systematic gaps due to OCP-Place were judged less wordlike than nonce verbs containing consonant pairs that are accidental gaps of equally low frequency. In these stimuli, the similarity of the nonce verbs to other roots in the Arabic lexicon was also controlled so that any effect of analogy to existing Arabic verbs would be equivalent for stimuli with OCP-Place violations and stimuli without OCP-Place violations. Sample stimuli with and without OCP-Place violations are given in (9). Since the influences of consonant frequency and analogy were controlled, the difference in acceptability of novel verbs with OCP-Place violations and novel verbs without OCP-Place violations could only be accounted for by positing a psychologically real OCP-Place constraint.

(9) /tasaba/
/tahafa/

In a second set of specially controlled nonce verbs, the same lexical factors were held constant while the similarity of consonant pairs that were OCP-Place violations was varied. Sample stimuli are given in (10). Ratings for this set of stimuli were sensitive to consonant pair similarity for the OCP-Place violations. This set of stimuli provides direct evidence that native Arabic speakers have implicitly learned a gradient OCP-Place constraint that reflects a similarity-based consonant co-occurrence restriction.

(10) /babaθa/ (identical)
/θabama/ (similar adjacent)
/baʃafa/ (similar nonadjacent)
/baʔada/ (nonhomorganic)

A second piece of evidence for the psychological reality of a gradient OCP-Place constraint comes from the influence of OCP-Place on the borrowing of lexical items into Maltese. Maltese is a historically Semitic language that has been heavily influenced by contact with Sicilian and Italian (Aquilina 1959). Maltese has a Semitic stratum that includes a non-concatenative morphological system of verb roots. However, Maltese also has an Italian stratum of verb stems that have not been analyzed into consonantal roots. Many words that are historically Italian have been incorporated into the Semitic stratum of Maltese, as evidenced by their productive use in the root-and-pattern system of morphology. Examples from Mifsud (1995) are shown in (11).

(11) /ziden/ 'to undo a knot in a fishing-line' < It. *snodare*
/dʒesses/ 'to chalk, plaster' < It. *gessare*
/tejjez/ 'to chop off s.o.'s head' < It. *tosare*

Among the incorporated roots, there appears to be an effect of OCP-Place. In other words, Italian words that violate OCP-Place are less likely to be incorporated into the Semitic stratum. We examined the consonant co-occurrence patterns in a corpus of verb borrowings compiled by Mifsud (1995). Since we claim below that OCP-Place effects are cross-linguistically common, we also examined consonant co-occurrence in a sample of comparable size of Italian verbs taken from a learner's dictionary (Colaneri 1992) to ensure that the co-occurrence patterns observed in the Maltese borrowings are not explained entirely by co-occurrence constraints that already exist in Italian. For the Maltese data, consonant co-occurrence was examined for adjacent consonant pairs in the verbal root, as in the analysis of Arabic. For the Italian data, consonant co-occurrence was examined over pairs of singleton consonants in the stem separated by a vowel.

Table VI shows observed and expected co-occurrence in the Maltese and Italian verb corpora for groups of identical consonant pairs and similar homorganic consonant pairs that are equivalent to the major classes of Arabic. The rate of co-occurrence for combinations of coronal stops and fricatives, and combinations of coronal obstruents and sonorants are given separately, since these groups are the most robust indicators of the influence of similarity on co-occurrence in Arabic. Note first that the co-occurrence patterns reflect an influence of similarity on both the Maltese borrowings and the Italian verb stems. However, the influence of OCP-Place is somewhat stronger and more like the Arabic pattern for the Maltese borrowings. The differences in strength of OCP-Place effects between the two corpora is significant (chi-square test over identical, similar homorganic, and coronal stop/fricative pairs, $\chi^2(2) = 9.0$, $p < 0.05$). This suggests that the strong OCP-Place constraint in the Semitic stratum of Maltese verbs has influenced which Italian verbs have been incorporated into the Semitic verb root system. Note, however, that the O/E for the borrowed roots is higher than what was found for equivalent Arabic roots. This suggests that OCP-Place had an influence on the borrowing of forms, but such borrowing was not blocked by the presence of an OCP-Place violation. Rather, phonotactically compatible forms were more likely to be borrowed than non-compatible forms. This influence can only be accounted for if the gradient patterns of consonant co-occurrence in the verbal roots of Maltese are a psychologically real component of the phonotactic knowledge of Maltese speakers.

4.2. *Similarity-Based Constraints Cross-Linguistically*

Phonotactic constraints on consonant place like those in Arabic are found in other Semitic languages (Bender and Fulass 1978; Buckley 1997; Greenberg 1950; Hayward and Hayward 1989; Koskinen 1964), as well as in Cambodian (Yip 1989), English (Berkley 1994a), French (Plenat 1996), Javanese (Mester 1986), Luo/Alur (Yip 1989), Ngbaka (Broe 1995), Pomo (Yip 1989), Ponapean (Yip 1989), Russian (Padgett 1995), Tsou (Wright 1996), and Yucatec Maya (McCarthy 1989). Formally related constraints

TABLE VI

Consonant co-occurrence in Maltese roots borrowed from Italian and in a sample of Italian verbs

	Maltese			Italian		
Consonant class	Obs	Exp	O/E	Obs	Exp	O/E
Identical	7	27.2	0.26	11	31.4	0.35
Similar Homorg	26	57.9	0.45	31	54.3	0.57
Cor Stop/Fric	13	16.7	0.78	33	28.8	1.15
Cor Obs/Son	111	80.8	1.37	76	66.7	1.14
Non-Homorg	206	182.7	1.13	271	250.2	1.08

are found for tone (Leben 1973; Goldsmith 1979; Odden 1986; Pierrehumbert and Beckman 1988) and for laryngeal features (Carré et al. 1995; Ito and Mester 1986; MacEachern 1999; Steriade 1982). Thus, the patterns we observe in Arabic can be found in a variety of unrelated languages. Additional implications for phonological theory are likely to come to light when these languages are studied in greater detail using the tools developed here.

There are examples in other languages where OCP-Place has been studied quantitatively. Padgett (1995) found a large number of exceptions to OCP-Place in a relatively small corpus of Russian, suggesting that the co-occurrence constraints in Russian are gradient. In several other cases, there is direct evidence for an influence of similarity on co-occurrence. Buckley (1997) replicated the results of Pierrehumbert (1993) for the Semitic language Tigrinya. The data on Maltese borrowings above are also suggestive of a similarity based OCP-Place constraint in both Maltese and Italian. Berkley (1994a, b, 2000) presents an extensive analysis of English that reveal gradient effects of similarity and distance on consonant co-occurrence. A study of consonant co-occurrence in Thai monosyllables from a learner's dictionary (Haas 1955) has also found gradient effects of similarity and distance on consonant co-occurrence (Frisch 2000b). For example, the rate of co-occurrence of similar homorganic consonants separated by a short vowel, as in /p^{h}óp/ 'to meet' (O/E = 0.53), is less than the rate of co-occurrence of similar homorganic consonants separated by a long vowel, as in /p^{h}âap/ 'picture, image' (O/E = 0.76). Also consistent with the similarity account, Thai has a gradient split between coronal obstruents and sonorants, with a much higher co-occurrence across the obstruent/sonorant contrast (O/E = 0.91) than within the obstruent or sonorant classes (O/E = 0.65).

There is another domain in which we speculate that similarity can play an explanatory role in capturing cross-linguistic generalizations. The very common process of vowel harmony forces all vowels in some domain to become more similar on one or more dimensions. In numerous African languages for example, all vowels in the word must have either an advanced or retracted tongue root (ATR harmony; see Archangeli and Pulleyblank 1989). Like the dissimilation of homorganic consonants in Arabic, many assimilations are sensitive to vowels that share a particular featural dimension. Cole and Trigo (1989) discuss a variety of cases of 'parasitic' vowel harmony. In Yawelmani, for example, a target vowel is forced to agree with a trigger vowel in rounding just in case it already agrees in height. These harmony processes are believed to be categorical, though they have not been examined quantitatively.

There are also harmony systems that are known to have non-absolute but statistically robust patterns of co-occurrence. Carré et al. (1995), employing an acoustically motivated classification of vowels based on perturbation theory, show that French exhibits a statistical tendency toward vowel harmony. Karlsson (1971) reports that in Finnish there exists a significant tendency to rounding harmony, in addition to the much studied and almost absolute front-back harmony. The front-back harmony controls the vocalic quality of suffixes in Finnish, while rounding harmony is confined to the stem. Like the experiments on Arabic described above, experiments on the psychological reality of harmony constraints in Finnish speakers suggest that perceptual similarity and acoustic salience play a role in determining how harmony is applied in novel derived words (Ringen and Heinamaki 1999).

Harmony may also occur among consonants, where it is also generally confined to an increase in similarity along one featural dimension (see Shaw 1991; Hansson 2001). We note here that strident consonants are especially prone to harmony: Chumash, Quechua, Kinyarwandi, and Navaho all exhibit forms of 'sibilant harmony' among the stridents /s/ and /ʃ/, but none of the other coronals. Stridents have the greatest degree of acoustic salience of any fricative (Shadle 1985), and the preponderance of sibilant harmony over other types of coronal harmony can be explained as an assimilation of features with the highest degree of perceived similarity. We wish to highlight the explanatory potential of similarity in all of these processes, and to suggest that a similarity-based description and quantitative analysis may reveal previously undiscovered patterns, as was found in the Arabic case. Some discussion of the issues involved in applying the natural classes similarity metric to consonant harmony can be found in Hansson (2001).

In general, the natural classes similarity metric we employed makes explicit the importance of dimensions of contrast in the segment inventory. While the distinctive features play a role in defining the natural classes, the classes crucially depend on the language particular contrasts available in the segment inventory. The same set of features applied to two different inventories defines two different natural class hierarchies. Other cross-linguistic differences in similarity effects may be found using the same similarity metric proposed here. Work by Padgett (2002) and Homer (1995) on harmony systems converges on analogous notions. Padgett's analysis of color harmony and Homer's analysis of nasal assimilation crucially depend on the contrastiveness of segmental targets, and not on the individual features involved. These works reflect a growing trend to consider the importance of contrast and perceptual features in explaining phonological patterns (e.g., Lindblom 1983; Flemming 1995; Silverman 1997). Traditional explanations of these processes as arbitrary formal operations involving feature spreading on symbolic representations are unable to account for the failure of spreading to apply when segmental contrasts would not be maintained.

4.3. *Explaining OCP-Place*

We view phonotactic acceptability, as reflected in the pattern of lexical items in Arabic, as a fundamentally gradient notion. The psychological category of phonotactically acceptable words that is in the minds of native speakers is grounded in a variety of phonological and lexical factors. However, it is important to note that extending our investigation and analysis to gradient data does not alter much of the fundamental nature of linguistic generalizations. Berent and Shimron (1997) and Frisch and Zawaydeh (2001) show that native speaker judgments of acceptability for novel Semitic roots crucially reflect the abstract natural classes defined by features. Our analysis of OCP-Place uses a similarity metric that is constrained to apply only to homorganic consonants. We propose to maintain the basic categorical structure of grammar, grounded in distinctive features and natural classes, but adopt a more psychologically realistic formulation of phonotactic knowledge.

In the O/E measure, phonological frequency plays an important role in determining baseline acceptability. The frequency of occurrence of Arabic consonants provides the Arabic learner with important information about what consonant combinations can be reasonably expected to occur. Positing a phonological constraint is warranted when observed co-occurrence deviates systematically from expected co-occurrence. The similarity-based OCP-Place constraint has an identifiable boundary between clearly acceptable and clearly unacceptable forms. The prototypically acceptable root in Arabic contains three consonants from different places of articulation. The prototypically unacceptable root contains clearly homorganic consonant pairs, such as identical consonant pairs. We claim that the Arabic consonant co-occurrence patterns are synchronically grounded in a gradient linguistic constraint, based on the perceived similarity of homorganic consonant pairs. Perceived similarity is a combination of the (paradigmatic) similarity formalized in section 3, conditioned by interference due to (syntagmatic) temporal distance. Consonant pairs that are highly similar are clearly homorganic to the language learner. Consonant pairs that are dissimilar are less obviously homorganic to the language learner, and thus are less likely to be considered violations of the OCP-Place constraint. We propose that the native speaker knows an abstract but gradient OCP-Place constraint ('Roots with repeated homorganic consonants are unusual'), based on generalization over the statistical patterns found in the lexicon. For example, the nonce verb /babaθa/ contains initial identical root segments. This nonce root has very low well-formedness, as it is of a pattern that is very rare and is in clear violation of OCP-Place. The nonce verb /baʃafa/ contains a non-adjacent non-homorganic stop-fricative pair. This form would have a low to moderate level of well-formedness, as analogous roots are attested but the labial pair still violates the broader generalization of OCP-Place.

The question remains, however, as to why repeated similar homorganic consonants are avoided in Arabic and other languages. We might further wonder why the constraint is so strong in Arabic, where many consonant pairs cannot co-occur at all, and much weaker in other languages such as Russian, English, and Thai. Pierrehumbert (1993) laid the groundwork for an explanation in the discussion of her similarity account. Berg (1998), Boersma (1998), and Frisch (1996, 2000a, in press) explicitly argue that repetition of similar consonants is difficult to process. This functional difficulty leads, diachronically, to a lexicon with few homorganic consonant pairs.

Berg (1998) claims that the repetition of consonants in Arabic root morphemes leads to potential confusion during the serialization of the segment sequence for the verb in language production. Berg further points out that this is particularly problematic in the case of Arabic, where consonant sequences in root morphemes could easily be confused due to the non-concatenative morphological system. Berg and Abd-El-Jawad (1996) discuss evidence that Arabic verb root consonants are unusually susceptible to certain kinds of speech errors involving consonant misordering, as in the following examples from Abd-El-Jawad and Abu-Salim (1987).

(12) /takriib/ for /takbiir/ ‘glorification’
/maraaʕiʃ/ for /maʃaaʕir/ ‘feelings’

Berg and Abd-El-Jawad claim that word-internal misordering errors are rare in English. They account for this language particular susceptibility for misordering by noting that abstract consonant root sequences are an important part of the morphological system of Arabic. In Arabic, the consonants of the root are a psychologically real entity distinct from the vowels, and it is at this morphological level that these misorderings take place in production. Prunet, Béland, and Idrissi (2000) found a similar language specific susceptibility to ordering errors in an Arabic speaking asphasic patient. Their patient was also a native speaker of French. While their patient frequently made errors in serial ordering of the consonants of Arabic roots, such errors were absent in French and in the ordering of Arabic consonants outside of the root in prefixes, suffixes, or fixed templatic positions.

Since segmental OCP effects are also found in languages that do not have non-concatenative morphology (e.g., Russian), it must be the case that the difficulty in serialization caused by repetition is a universal of language processing, though the effects of the constraint may be weaker or stronger depending on the other aspects of a language’s phonology and morphology. In fact, there is ample evidence from psycholinguistic experiments with English speakers that reveal a difficulty in processing repetition in production, perception, and working memory tasks (see Frisch, in press, for a thorough overview). For example, speech error rates are increased in utterances in which segments are repeated in proximity to one another (Dell 1984). Segmental repetition also slows the overall production rate (Sevald and Dell 1994). In perception tasks, two rapidly repeated identical stimuli are often not detected as distinct, and only a single token is reported heard (Miller and MacKay 1994). Therefore, we assume that repetition is more difficult to process than non-repetition in all languages. In the case of Arabic, where the root consonants are stored as a distinct level of the lexical representation, the processing difficulties will be especially great, as the root consonants are in close proximity at that level.

Boersma (1998) primarily focuses on a perceptual difficulty with repetition. In particular, he examines the case of parsing a sequence of segments where there is immediate repetition of a segment. In this case, it is difficult to recover the independent existence of both segments. Boersma argues that phonological processes such as epenthesis, dissimilation, and blocking of vowel deletion exist to avoid this processing difficulty. There is reason to believe that perceptual problems with processing repetition may apply for non-local repetition as well. In speech perception, the rate of information transmission is very high and the identification of segments must be very rapid. However, the speech perception module has to deal with many cases of missing information or ambiguity. One possible aid in disambiguation is higher order information, such as the knowledge of the identity of a word. This knowledge supplies information about the segments in that word. Thus, the perceptual system does not immediately make a firm commitment to any particular percept of a segment, as additional information might cause a change in the identity of missing or ambiguous segments (see Klatt 1989 for discussion). With delayed decisions about segmental identity, the repetition of similar segments within a word may result in blending of perceptual traces and consequently result in the misperception of segments.

The lexical roots of Arabic directly (quantitatively) reflect the proposed phonetic explanation. Repeated similar homorganic consonants are avoided to the extent that they are similar. In this way, statistical analysis of the lexicon provides a novel type of evidence for functionally motivated constraints (see Hawkins 1994 for similar arguments at the syntactic level). The lexicon quantitatively reflects the gradient nature of the underlying phonetic motivation (cf. Hayes 1999). Statistical patterns in the lexicon can be seen as the result of the diachronic influence of the processing constraint against repetition. Over time, functional pressures on the language have shaped the lexicon that is to be acquired by successive generations of speakers. These functional pressures influence borrowing, the creation of nonce forms, and the loss of lexical items. The grounding of the similarity-based OCP-Place constraint in language processing accounts for both the gradient nature of the constraint and the common occurrence of these constraints cross-linguistically. Despite the diachronic origin of the dissimilation patterns, native Arabic speakers have acquired the OCP-Place constraint and so it must be considered a part of the synchronic linguistic knowledge of the speakers.

5. Discussion

We have argued that the degree of co-occurrence restriction between consonants in the Arabic verbal roots depends on place of articulation, manner, and voicing features, as well as the distance between consonant pairs. The autosegmental OCP account used separate rules for what were seen as two different effects of two different strengths, one on identical consonants (total OCP) and one on homorganic consonants (OCP-Place). OCP-Place was known to be manner sensitive for coronals, for which a specific sub-rule was proposed. We have shown the effects of manner are much more widespread than previously noted, as some manner effects are found within every place class. Further, the degree of the manner effect varies from class to class. The strongest manner effect is found for coronals, as coronal obstruents and sonorants frequently co-occur. A weaker manner effect was found for dorsals, gutturals, and labials. In a categorical OCP account, each of these cases would require an additional sub-rule. Additional sub-cases are required for voicing features as well. This results in a combinatorial explosion of seemingly unrelated sub-cases, and no a priori predictions of the strength of the effect for each of these cases. By contrast, the similarity avoidance account provides an alternative that incorporates all of these sub-cases under a single generalization. Different degrees of well-formedness follow from the degree of similarity between consonants. Similarity is contrast sensitive and so is dependent upon the size and structure of the inventory. Larger classes contain item pairs with lower similarity values than analogous pairs in smaller classes. In addition, any feature is potentially relevant to similarity, and so many sub-patterns of consonant co-occurrence are predicted.

Optimality Theoretic approaches to the segmental OCP using conjoined markedness constraints have attempted to capture similarity effects by positing a constraint for each distinct sub-case of feature co-occurrence (e.g., $^*[\text{labial}]^2$, $^*([\text{labial}]\ \&\ [-\text{son}])^2$, $^*([\text{labial}]\ \&\ [-\text{son}]\ \&\ [+\text{cont}])^2$, see Alderete 1997 and MacEachern 1999). Since a violation of $^*[\text{labial}]^2$ implies a violation of $^*([\text{labial}]\ \&\ [-\text{son}])^2$, there is a natural ranking of these constraints in a similarity hierarchy ($^*([\text{labial}]\ \&\ [-\text{son}]\ \&\ [+\text{cont}])^2 \gg {}^*([\text{labial}]\ \&\ [-\text{son}])^2 \gg {}^*[\text{labial}]^2$). This ranking is natural because more specific constraints must outrank more general ones in order to be expressed in the outcomes (see discussion of the Elsewhere Condition in Prince and Smolensky 1993). However, applying such accounts to the Arabic case fails to predict the ranking of complex constraints that involve different feature combinations resulting in the same degree of similarity. It also leaves the effect of inventory size on similarity unexplained. By failing to define a similarity metric, it misses a generalization about the nature of the effects. Further, the OT account predicts that there would be a similarity threshold below which all forms are categorically impossible, and above which all forms are categorically possible, based on the placement of faithfulness constraints with respect to the similarity constraint family.

Recently, several OT researchers have taken on issues in the variationist and functionalist literature. They have extended the classical OT model by adding stochastic ranking of constraints to account for quantitative variation between forms (Anttila 1998; Boersma 1998; Boersma and Hayes 2001; Hayes and MacEachern 1998). However, their approaches retain the fundamental evaluation procedure of OT, with strict constraint dominance determining the winning candidate in a given evaluation of the candidate set. Quantitative variation can be accounted for if some rankings produce one variant while other rankings produce another variant for different trials evaluating the same candidate set. Boersma and Hayes (2001) demonstrate that a learning algorithm using stochastic ranking can match probabilities for variants quite accurately. They extend the same mechanism that was used to model variation to model instances of graded well-formedness. In their model, graded well-formedness is a function of the probability of the output of a form over a large number of trials that use stochastically ranked constraints to evaluate that form in comparison to the alternative candidates. They propose that the form is judged as highly well-formed if it is the best candidate in most of the trials, and it is judged as slightly well-formed if it is the best candidate in just a few trials. However, the constraints and their interaction in any particular evaluation are still categorical. Frequencies and graded well-formedness result from variation across individual evaluations that individually produce categorical results. Our approach is much simpler, in that we claim that constraints and their interaction can be gradient, with graded well-formedness determined directly. Using a gradient constraint provides a straightforward account of graded phonotactic well-formedness in the lexicon.

The patterns in the Arabic roots do not reflect variation in the output for individual roots, but rather involve consistent realization for any particular root. In order to derive consistent realization for a particular root, stochastic OT accounts require some additional meta-evaluation mechanism that can separate the instances where evaluation is conducted to determine well-formedness from the instances where evaluation is conducted to produce an output from an input. In instances where an output is produced, the constraint ranking must be limited to one of the set of rankings that allows the root to be well-formed (Berkley 1994b). But such a fixing of constraints must not be allowed to happen in cases where free variation occurs. Boersma and Hayes (2001) use the same formal mechanism to derive ordinary outputs, cases of variation, and cases of graded well-formedness, and so cannot explain quantitative lexical patterning without predicting variation for those patterns.

Additional evidence for the gradience of constraints and constraint interaction can be seen in the asymmetries in consonant co-occurrence discussed by Elmedlaoui (1995). These gradient sonority-based asymmetries interact with the similarity-based OCP. Also, Hay, Pierrehumbert, and Beckman (2003) found an analogous case in which OCP-Place in English interacts with the homorganic NC constraint. In both cases, the constraint interaction is not one where either constraint dominates the determination of well-formedness. Rather, both constraints act cumulatively. Frisch (2000a) develops a formal model of gradient constraint interaction for multiple OCP-Place violations in the Arabic roots using the stochastic constraint model.

We propose that the OCP-Place constraint learned by Arabic speakers is based on abstraction over the observed set of roots in the lexicon. The fact that consistent outputs are realized for any particular root follows straightforwardly from lexical memory and morphological competence. It is the total lexicon of roots, and not the output corresponding to any particular input root, that reflects the influence of OCP-Place. The patterns of consonant co-occurrence in Arabic can only be explained if the lexicon itself plays a more important role in determining the phonology of the language learner. Since we claim that OCP-Place patterns are functionally grounded in similarity avoidance, similarity avoidance itself does not directly play a role in the learner's construction of a grammar (*contra* Hayes 1999). Rather, similarity avoidance shapes the set of lexical items that the learner encounters. We claim that lexical items that avoid repetition will be easier to process, and so will be favored in acquisition, lexical borrowing, coining novel forms, and in active usage. Once a similarity avoidance pattern becomes established, it will be further reinforced by the grammars of the speakers that learn the pattern, since grammar influences borrowing and novel word formation as well. In this way, relatively weak functional forces can result in grammars with strong constraints that avoid functionally weak forms. Overall, we conclude that this 'feedback loop' connecting the details of speech processing with the phonological grammar has two effects: the phonological grammar and the lexicon conform to one another, and the grammar contains constraints that are functionally grounded.

We do not claim, however, that constraints cannot be categorical, or that constraint interaction never involves categorical dominance. As MacEachern (1999) points out, the case of repeated final consonants in the Arabic roots provides a ready example. The morphological abstraction of these roots as biconsonantal supercedes the preference of OCP-Place to avoid repetition, and the resulting distribution of forms does not follow from the cumulative interaction of the OCP with some other constraint licensing final repetition. Frisch and Zawaydeh (2001) found forms with repeated final consonants to be judged as highly well-formed by Arabic speakers (see also Berent and Shimron 1997). The formalism of Frisch (1996, 2000a) and Frisch, Broe, and Pierrehumbert (1997) provides a way for individually categorical constraints to arise as the mathematical limit of sharpening gradient constraints, but does not address categorical constraint interaction. The interaction of gradient constraints or of categorical constraints with gradient constraints is currently an outstanding unsolved problem.

The inadequacy of the autosegmental OCP to account for the effects of similarity and distance on co-occurrence shows that the domain over which phonotactic generalizations can apply is more complex than what is predicted by tier segregation and feature geometry. Phonotactic generalizations can be drawn over truly non-local domains, to non-adjacent objects (see Berkley 2000 for additional evidence from English and Latin). Segmental OCP constraints are not unique in this respect. For example, non-local OCP-Tone effects have been shown to be sensitive to prosodic and morphological structure (Odden 1986; Pierrehumbert and Beckman 1988). More generally, since gradient OCP effects can be found non-locally, through intervening material and phonological and morphological boundaries, other gradient phonotactic generalizations are likely to exist. A growing body of evidence demonstrates that very young infants begin to form such generalizations well before they are capable of using language to receive or convey semantic information (Jusczyk 1997 inter alia; Saffran et al. 1998). The increasing awareness in the literature of the psychological reality of gradient phonotactic generalizations indicates that the traditional generative model, in which the lexicon plays no significant role in the active phonology, is incorrect. However, these phonotactic generalizations are not made over visceral perceptual patterns either. Phonotactic generalizations depend on segmental and prosodic structure (Coleman and Pierrehumbert 1998; Frisch et al. 2000; Treiman et al. 2000; Vitevitch et al. 1997), and so they in turn are dependent upon other aspects of the phonology. We conclude that the lexicon provides a rich domain over which phonological generalizations can be constructed at a number of levels of representation.

6. Conclusion

Gradient constraints, like the similarity based OCP-Place constraint, have been proposed for cases of phonetic implementation, which is often divorced from phonology proper (e.g., Keating 1984; Pierrehumbert and Beckman 1988). The phonetics-phonology division was originally motivated by a desire to separate the symbolic phonological system from the probabilistic and gradient nature of speech. We conclude that the Arabic speaker's knowledge of OCP-Place reflects implicit linguistic knowledge about the possible words in a language, such that gradient phenomena must be incorporated within phonology proper. It has also been found that rules of phonetic implementation are language specific, which undermines the existence of a dividing line between phonology and phonetics. We believe that the unification of phonological and phonetic knowledge into a single system allows gradient effects within the two domains to be accounted for with analogous mechanisms (see Pierrehumbert, forthcoming). Since many phonological processes are phonetically motivated, such a synthesis is inevitable as our ability to explain phonological phenomena grows (Hayes 1999). This synthesis raises new challenges for phonological theory, as the incorporation of phonetic and lexical information into the grammar requires a phonological formalism that can integrate discrete and continuous patterns. The phonotactic patterns of Arabic provide evidence that a quantitative phonotactic component based on lexical patterns is a core element of this grammar.

REFERENCES

Abd-El-Jawad, Hassan and Issam Abu-Salim. 1987. 'Slips of the Tongue in Arabic and Their Theoretical Implications', *Language Sciences* **9**, 145–171.

Alderete, John. 1997. 'Dissimilation as Local Conjunction', in K. Kiyomi (ed.), *Proceedings of the North East Linguistics Society* 27, pp. 17–32.

Anttila, Arto. 1998. 'Deriving Variation from Grammar', in Frans Hinskens, Roeland Van Hout and W. Leo Wetzels (eds.), *Variation, Change, and Phonological Theory*, John Benjamins, Amsterdam, pp. 35–68.

Aquilina, Joseph. 1959. *The Structure of Maltese: A Study of Mixed Grammar and Vocabulary*, Royal University of Malta, Valetta.

Archangeli, Diana and Douglas Pulleyblank. 1989. 'Yoruba Vowel Harmony', *Linguistic Inquiry* **20**, 173–217.

Bachra, Bernard N. 2000. *The Phonological Structure of the Verbal Roots in Arabic and Hebrew*, Brill, Leiden.

Bailey, Todd and Ulrike Hahn. 2001. 'Determinants of Wordlikeness: Phonotactics or Lexical Neighborhoods?', *Journal of Memory and Language* **4**, 568–591.

Bender, M. Lionel and Hailu Fulass. 1978. *Amharic Verb Morphology*, Michigan State University African Studies Center, East Lansing.

Berent, Iris and Joseph Shimron. 1997. 'The Representation of Hebrew Words: Evidence from the Obligatory Contour Principle', *Cognition* **64**, 39–72.

Berg, Thomas. 1998. *Linguistic Structure and Change: An Explanation from Language Processing*, Clarendon Press, Oxford.

Berg, Thomas and Hassan Abd-El-Jawad. 1996. 'The Unfolding of Suprasegmental Representations: A Cross-linguistic Perspective', *Journal of Linguistics* **32**, 291–324.

Berkley, Deborah Milam. 1994a. 'Variability in Obligatory Contour Principle Effects', in *Papers from the 30th Regional Meeting of the Chicago Linguistics Society*, University of Chicago Press, Chicago, pp. 1–12.

Berkley, Deborah Milam. 1994b. 'The OCP and Gradient Data', *Studies in the Linguistic Sciences* **24**(1/2), 59–72.

Berkley, Deborah Milam. 2000. 'Gradient OCP Effects', unpublished Ph.D. dissertation, Northwestern University.

Boersma, Paul. 1998. *Functional Phonology*, Holland Academic Graphics, The Hague.

Boersma, Paul and Bruce Hayes. 2001. 'Empirical Tests of the Gradual Learning Algorithm', *Linguistic Inquiry* **32**, 45–86.

Broe, Michael. 1993. *Specification Theory: The Treatment of Redundancy in Generative Phonology*, unpublished Ph.D. dissertation, University of Edinburgh.

Broe, Michael. 1995. 'Specification Theory and Ngbaka Co-occurrence Constraints', paper presented at the Mid-Continental Workshop on Phonology, Columbus, Ohio.

Broe, Michael. 1997. 'Stochastic Constraints in Phonological Analysis', paper presented at Indiana University, Bloomington, Indiana.

van den Broeke, Marcel and Louis Goldstein. 1980. 'Consonant Features in Speech Errors', in Victoria Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*, Academic Press, New York, pp. 47–65.

Buckley, Eugene. 1997. 'Tigrinya Root Consonants and the OCP', *Penn Working Papers in Linguistics* **4.3**, 19–51.

Carré, R., M. Bourdeau and J-T. Tubach. 1995. 'Vowel-vowel Production: The Distinctive Region Model DRM and Vowel Harmony', *Phonetica* **52**, 205–214.

Colaneri, John. 1992. *501 Italian Verbs*, Barron's, Hauppauge, New York.

Cole, Jennifer and Loren Trigo. 1989. 'Parasitic Harmony', in H. van der Hulst and N. Smith (eds.), *Features, Segmental Structure and Harmony Processes*, Foris, Dordrecht, pp. 19–38.

Coleman, John and Janet B. Pierrehumbert. 1998. 'Stochastic Phonological Grammars and Acceptability', in *Proceedings of the Third Meeting of the ACL Special Interest Group in Computational Phonology*, Association for Computational Linguistics, Somerset, New Jersey, pp. 49–56.

Cowan, J. 1979. *Hans Wehr: A Dictionary of Modern Written Arabic*, Otto Harrasowitz, Wiesbaden, Germany.

Dell, Gary S. 1984. 'Representation of Serial Order in Speech: Evidence from the Repeated Phoneme Effect in Speech Errors', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **10**, 222–233.

Devore, J. L. 1987. *Probability and Statistics for Engineering and the Sciences*, 2nd ed., Brooks/Cole, Monterey, California.

Elmedlaoui, Mohamed. 1995. 'Géométrie des restrictions de co-occurrence de traits en sémitique et en berbère: Synchronie et diachronie', *Canadian Journal of Linguistics* **40**, 39–76.

Flemming, Edward. 1995. *Auditory Representations in Phonology*, unpublished Ph.D. dissertation, UCLA.

Frisch, Stefan A. 1996. *Similarity and Frequency in Phonology*, unpublished Ph.D. dissertation, Northwestern University.

Frisch, Stefan A. In press. 'Language Processing and Segmental OCP Effects', in B. Hayes, R. Kirchner, and D. Steriade (eds.), *Phonetically-based Phonology*, Cambridge University Press, Cambridge.

Frisch, Stefan A. 2000a. 'Temporally Organized Representations as Phonological Units', in M. B. Broe and J. B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, Cambridge University Press, Cambridge, pp. 283–298.

Frisch, Stefan A. 2000b. 'Competence and Performance in the Phonological Lexicon', paper presented at the University of California, Irvine.

Frisch, Stefan A., Michael B. Broe and Janet B. Pierrehumbert. 1997. 'Similarity and Phonotactics in Arabic', unpublished manuscript, Indiana University, available from the Rutgers Optimality Archive (http://roa.rutgers.edu/).

Frisch, Stefan A., Nathan R. Large and David B. Pisoni. 2000. 'Perception of Wordlikeness: Effects of Segment Probability and Length on the Processing of Nonwords', *Journal of Memory and Language* **42**, 481–496.

Frisch, Stefan A. and Bushra Zawaydeh. 2001. 'The Psychological Reality of OCP-Place in Arabic', *Language* **77**, 91–106.

Gafos, Adamantios. 1998. 'Eliminating Long-distance Consonant Spreading', *Natural Language and Linguistic Theory* **16**, 223–278.

Gafos, Adamantios. 2001. 'The Initial State and Verbal Stems in Arabic', unpublished manuscript, Utrecht Institute of Linguistics, OTS.

Goldsmith, John. 1979. *Autosegmental Phonology*, Garland, New York.

Greenberg, Joseph H. 1950. 'The Patterning of Root Morphemes in Semitic', *Word* **5**, 162–181.

Greenberg, Joseph H. and John J. Jenkins. 1964. 'Studies in the Psychological Correlates of the Sound System of American English', *Word* **20**, 157–177.

Haas, Mary R. 1955. *Thai Vocabulary*, American Council of Learned Societies, Washington, D.C.

Hansson, Gunnar. 2001. *Theoretical and Typological Issues in Consonant Harmony*, unpublished Ph.D. dissertation, University of California, Berkeley.

Hawkins, John A. 1994. *A Performance Theory of Word Order and Constituency*, Cambridge University Press, Cambridge.

Hay, Jennifer, Janet B. Pierrehumbert and Mary Beckman. 2003. 'Speech Perception, Wellformedness, and the Statistics of the Lexicon', in *Papers in Laboratory Phonology VI*, Cambridge, Cambridge University Press, pp. 58–74.

Hayes, Bruce. 1999. 'Phonetically Driven Phonology: The Role of Optimality Theory and Inductive Grounding', in M. Darnell, E. Moravcsik, F. Newmeyer, M. Noonan, and K. Wheatley (eds.), *Functionalism and Formalism in Linguistics*, John Benjamins, Amsterdam, pp. 243–285.

Hayes, Bruce and Margaret MacEachern. 1998. 'Quatrain Form in English Folk Verse', *Language* **74**, 473–507.

Hayward, K. and R. Hayward. 1989. '"Guttural": Arguments for a New Distinctive Feature', *Transactions of the Philological Society* **87**, 179–193.

Homer, Molly. 1995. 'Segment Sets Account for Opacity in Applecross Gaelic', paper presented at the Midcontinental Workshop in Phonology, Ohio State University, Columbus, Ohio.

Ito, Junko and R. Armin Mester. 1986. 'The Phonology of Voicing in Japanese', *Linguistic Inquiry* **17**, 49–73.

Jusczyk, Peter. 1997. *The Discovery of Spoken Language*, MIT Press, Cambridge.

Karlsson, F. 1971. 'Finskans rotmorfemstruktur: en generativ beskrivning', *Turun Yliopston Fonetiikan Laitoksen Julkaisuja* No. 10, Turku.

Keating, Patricia. 1984. 'Phonetic and Phonological Representation of Stop Consonant Voicing', *Language* **60**, 286–319.

Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*, Blackwell, Oxford.

Kessler, Brett and Rebecca Treiman. 1997. 'Syllable Structure and the Distribution of Phonemes in English Syllables', *Journal of Memory and Language* **37**, 295–311.

Kirchner, Robert. 1997. 'Contrastiveness and Faithfulness', *Phonology* **14**, 83–111.

Klatt, Dennis H. 1989. 'Review of Selected Models of Speech Perception', in W. Marslen-Wilson (ed.), *Lexical Representation and Process*, MIT Press, Cambridge, pp. 169–226.

Koskinen, Kalevi. 1964. 'Kompatibilitat in den dreikonsonantigen hebraishen wurzeln', *Zeitschrift der Deutschen Morgenlandischen Gesellschaft* **144**, 16–58.

Leben, William. 1973. *Suprasegmental Phonology*, unpublished Ph.D. dissertation, MIT.
Liberman, Mark and Janet B. Pierrehumbert. 1984. 'Intonational Invariance under Changes in Pitch Range and Length', in M. Aronoff and R. Oehrle (eds.), *Language Sound Structure*, MIT Press, Cambridge, Massachusetts, pp. 157–233.
Lightner, Theodore. 1973. 'Against Morpheme Structure Rules and Other Things', in M. Kenstowicz and C. Kisseberth (eds.), *Issues in Phonological Theory*, Mouton, The Hague, pp. 53–60.
Lindblom, Bjorn. 1983. 'Economy of Speech Gestures', in P. MacNeilage (ed.), *The Production of Speech*, Springer-Verlag, New York, pp. 217–245.
MacEachern, Margaret. 1999. *Laryngeal Co-occurrence Restrictions*, Garland, New York.
McCarthy, John J. 1979. *Formal Problems in Semitic Phonology and Morphology*, Garland, New York.
McCarthy, John J. 1986. 'OCP Effects: Gemination and Antigemination', *Linguistic Inquiry* **17**, 207–263.
McCarthy, John J. 1988. 'Feature Geometry and Dependency: A Review', *Phonetica* **43**, 84–108.
McCarthy, John J. 1989. 'Linear Order and Phonological Representation', *Linguistic Inquiry* **20**, 71–100.
McCarthy, John J. 1994. 'The Phonetics and Phonology of Semitic Pharyngeals', in P. Keating (ed.), *Papers in Laboratory Phonology III*, Cambridge University Press, Cambridge, pp. 191–283.
Medin, Douglas L. and M. M. Shaeffer. 1978. 'A Context Theory of Classification Learning', *Psychological Review* **85**, 207–238.
Mester, R. Armin. 1986. *Studies in Tier Structure*, unpublished Ph.D. dissertation, University of Massachusetts, Amherst.
Mifsud, Manwell. 1995. *Loan Verbs in Maltese: A Descriptive and Comparative Study*, E. J. Brill, Leiden.
Miller, Michelle D. and Donald G. MacKay. 1994. 'Repetition Deafness: Repeated Words in Computer-compressed Speech Are Difficult to Encode and Recall', *Psychological Science* **5**, 47–51.
Nosofsky, Robert M. 1986. 'Attention, Similarity, and the Identification-categorization Relationship', *Journal of Experimental Psychology: General* **115**, 39–57.
Odden, David. 1986. 'On the Obligatory Contour Principle', *Language* **62**, 353–383.
Padgett, Jaye. 1995. *Stricture in Feature Geometry*, CSLI Publications, Stanford, California.
Padgett, Jaye. 2002. 'Feature Classes in Phonology', *Language* **78**, 81–110.
Pierrehumbert, Janet B. 1993. 'Dissimilarity in the Arabic Verbal Roots', in *Proceedings of the North East Linguistics Society* **23**, pp. 367–381.
Pierrehumbert, Janet B. 1994, 'Syllable Structure and Word Structure', in P. Keating (ed.), *Papers in Laboratory Phonology III*, Cambridge University Press, Cambridge, pp. 168–190.
Pierrehumbert, Janet B. 1999. 'Formalizing Functionalism', in M. Darnell, E. Moravscsik, F. Newmeyer, M. Noonan and K. Wheatley (eds.), *Formalism and Functionalism in Linguistics*, John Benjamins, Amsterdam, vol. 1, pp. 287–305.
Pierrehumbert, Janet B. Forthcoming. 'The Phonetic Grounding of Phonology', *Bulletin de Communication Parleé*.
Pierrehumbert, Janet B. and Mary Beckman. 1988. *Japanese Tone Structure*, MIT Press, Cambridge.

Plenat, Marc. 1996. 'De l'interaction des contraints: une etude de case', in J. Durand and B. Laks (eds.), *Current Trends in Phonology: Models and Methods*, University of Salford Press, Salford, pp. 585–617.

Prince, Alan and Paul Smolensky. 1993. *Optimality Theory*, Rutgers University Center for Cognitive Science Technical Report 2, Rutgers University, New Brunswick, New Jersey.

Prunet, Jean-François, René Béland and Ali Idrissi. 2000. 'The Mental Representation of Semitic Words', *Linguistic Inquiry* **31**, 609–648.

Repp, Bruno. 1984. 'Categorical Perception: Issues, Methods, Findings', in N. J. Lass (ed.), *Speech and Language: Advances in Basic Research and Practice*, vol. 9, Academic Press, New York, pp. 243–335.

Ringen, Catherine and Orvokki Heinamaki. 1999. 'Variation in Finnish Vowel Harmony: An OT Account', *Natural Language and Linguistic Theory* **17**, 303–337.

Rose, Sharon. 2000. 'Rethinking Geminates, Long-distance Geminates and the OCP', *Linguistic Inquiry* **31**, 85–122.

Saffran, Jenny R., Richard N. Aslin and Elissa L. Newport. 1998. 'Statistical Learning by 8-month-old Infants', *Science* **274**, 1926–1928.

Sevald, Christine and Gary S. Dell. 1994. 'The Sequential Cueing Effect in Speech Production', *Cognition* **53**, 91–127.

Shadle, Christine. 1985. *The Acoustics of Fricative Consonants*, unpublished Ph.D. dissertation, MIT.

Shaw, Patricia. 1991. 'Consonant Harmony Systems', in C. Paradis and J.-F. Prunet (eds.), *Phonetics and Phonology 2: The Special Status of Coronals*, Academic Press, San Diego, California, pp. 125–157.

Silverman, Daniel. 1997. *Phasing and Recoverability*, Garland, New York.

Stanley, Richard. 1967. 'Redundancy Rules in Phonology', *Language* **43**, 393–436.

Steriade, Donca. 1982. *Greek Prosodies and the Nature of Syllabification*, unpublished Ph.D. dissertation, MIT.

Steriade, Donca. 1987. 'Redundant Values', in *Papers from the 23rd Regional Meeting of the Chicago Linguistics Society*, University of Chicago Press, Chicago, pp. 339–362.

Steriade, Donca. 1995. 'Underspecification and Markedness', in J. Goldsmith (ed.), *The Handbook of Phonological Theory*, Blackwell, Oxford, pp. 114–174.

Treiman Rebecca, Brett Kessler, Stephanie Knewasser, Ruth Tincoff and Michael Bowman. 2000. 'Adults' Sensitivity to Phonotactic Probabilities', in M. B. Broe and J. B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, Cambridge University Press, Cambridge, pp. 269–282.

Tversky, Amos. 1977. 'Features of Similarity', *Psychological Review* **84**, 327–352.

Vitevitch, Michael S., Paul A. Luce, Jan Charles-Luce and David Kemmerer. 1997. 'Phonotactics and Syllable Stress: Implications for the Processing of Nonsense Words', *Language and Speech* **40**, 47–62.

Vitevitch, Michael S., Paul A. Luce, David B. Pisoni and Edward T. Auer. 1999. 'Phonotactics, Neighborhood Activation and Lexical Access for Spoken Words', *Brain and Language* **68**, 306–311.

Wright, Richard. 1996. *Consonant Clusters and Cue Preservation in Tsou*, unpublished Ph.D. dissertation, UCLA.

Yip, Moira. 1989. 'Feature Geometry and Co-occurrence Restrictions', *Phonology* **6**, 349–374.

Received 17 September 1997
Revised 22 July 2002

Stefan A. Frisch
Department of Communication
Sciences and Disorders
University of South Florida
4202 E. Fowler Ave, PCD1017
Tampa, FL 33620
USA
<frisch@chuma1.cas.usf.edu>

Janet B. Pierrehumbert
Department of Linguistics
Northwestern University
2016 Sheridan Road
Evanston, IL 60201
USA
<jbp@northwestern.edu>

Michael B. Broe
Department of Linguistics
The Ohio State University
222 Oxley Hall
1712 Neil Avenue
Columbus, OH
USA
<mbroe@ling.ohio-state.edu>