The psychological reality of an abstract consonant dissimilation constraint is demonstrated in an experiment with native speakers of Jordanian Arabic. In this experiment, novel verbs containing constraint violations and those without violations were presented orthographically for judgments of well-formedness. Native speaker well-formedness judgments reflected knowledge of the phonotactic constraint. Systematic gaps were rated much less wordlike than accidental gaps that were equivalent in their lexical characteristics. Judgments for novel verbs containing constraint violations were also gradiently influenced by consonant pair similarity. The experimental study supports previous dictionary-based phonotactic analyses that propose that the native speaker’s knowledge of consonant cooccurrence constraints in Arabic is based on emergent generalizations over the lexical items in an abstract root lexicon.*

The phonotactic constraints of a language define the set of possible words in that language. Typically, phonotactic constraints are restrictions on the possible consonant clusters in a language or conditions on the minimal prosodic unit that can be considered a word. Such constraints undoubtedly have an impact on productive word formation and foreign word accommodation, and thus these constraints are assumed to be active in the synchronic grammar. In the case of the Arabic dialects, there have been a number of studies (based on the analysis of a dictionary of Standard Arabic as an approximation to the mental lexicon) that claim that there is a phonotactic restriction on possible consonant sequences within the abstract verbal root (Greenberg 1950, McCarthy 1986, 1988, 1994, Pierrehumbert 1993, Frisch et al. 1997). The Arabic case is of particular interest as the root consonants do not usually appear adjacent to one another in the words of the language. Arabic has a root-and-pattern morphology, where consonants from the verbal root are interleaved with vowels in a manner prescribed by a lexical template. Due to the abstractness of lexical roots, an investigation of the psychological reality of consonant cooccurrence constraints defined over root consonants provides a strong test of abstract generative theories of competence.

In generative phonological theory (e.g. Kenstowicz 1994, Prince & Smolensky 1993) phonotactic constraints are typically categorical statements of well-formedness. Recent analyses of Arabic phonotactics, however, have found systematic quantitative regularities that cannot be explained by categorical constraints (McCarthy 1994, Pierrehumbert 1993, Frisch et al. 1997). But these analyses have been based on statistical studies using a dictionary as an approximation of the knowledge of lexical items possessed by native speakers. Thus, it is additionally unclear whether the detailed quantitative patterns in the lexicon are a part of the grammars of native speakers. It may be that quantitative lexical patterns are regularized and encoded as categorical in the grammar. Alternately, it could be the case that quantitative phonotactic constraints have no effect on the grammar at all. The statistical lexical patterns may nonetheless affect productive word formation or metalinguistic judgments of well-formedness via analogical pro-

* Thanks to the audiences at MCWOP 3 and LSA 1999 for questions and general discussion of this work. Thanks to Sharon Inkelas, Kie Zuraw, and the Language referees for helpful comments on an earlier draft of this article. This research was supported by NIH Training Grant DC00012 to Indiana University, an Indiana University Grant-in-Aid of Research, and the Language Learning Visiting Research Assistant Professorship at the University of Michigan.
cesses. In this case, we might say that these patterns are part of a native speaker’s synchronic knowledge, but they are external to the grammar.

The potential for lexical statistics to influence well-formedness judgments was taken into account in the design of stimuli for an experiment on native speaker knowledge of Arabic phonotactic patterns. Lexical statistics and the presence of phonotactic constraint violations were carefully manipulated in an attempt to determine the relative importance of these factors in speakers’ well-formedness judgments for novel roots. By controlling lexical factors, we obtain unambiguous evidence for the psychological reality of abstract, gradient phonotactic constraints in Arabic. Our results support recent studies of Arabic phonotactic patterns based on statistical analyses of dictionary corpora. Further, we demonstrate that these constraints must be abstract, and not merely a reflection of superficial lexical statistics or analogical comparison of relatively unanalyzed surface forms (e.g. as in the analogy-based grammar of Skousen 1992, see Eddington 2000). Due to the quantitative nature of this abstract constraint, the phonotactic knowledge employed by native speakers of Arabic cannot be described in current generative theory. Thus, we find phonotactic knowledge is more fine grained than generative theory allows, and more abstract than analogical word comparison would predict. We propose that linguistic competence includes emergent phonotactic constraints that are derived from qualitative and quantitative generalizations over the lexicon.

1. PHONOTACTIC CONSTRAINTS AND JUDGMENTS OF WELL-FORMEDNESS. Semitic languages have a nonconcatenative root-and-pattern morphology, in which the consonants and the vowels in the inflected verb can be analyzed as distinct morphemes. This situation provides a rationale for segregating the consonants and vowels into separate autosegmental tiers (McCarthy 1986), as in kutib ‘was written’ in 1. The consonantal root for this verb is /ktb/.

(1) consonant tier: k t b
   skeletal tier: C V C V C
   vowel tier: u i

Berg and Abd-El-Jawad (1996) present evidence from a crosslinguistic study of speech errors that supports the psychological reality of abstract consonantal roots. In English and German, speech errors involving interactions between consonants obey a SYLLABLE POSITION CONSTRAINT, such that onsets interact with onsets and codas interact with codas. Berg and Abd-El-Jawad examined speech errors in Jordanian Arabic and found that root consonants in different (surface) syllabic positions frequently interact. The fact that Arabic speech errors violate the syllable position constraint can be explained if errors sometimes occur at the level of the abstract root where consonants are not bound to specific syllable positions. Prunet, Béland and Idrissi (1999) found similar patterns in a study of production errors by a bilingual Lebanese Arabic and French speaking adult with aphasia. This patient was diagnosed as a deep dyslexic in both languages and he had particular difficulty in producing the correct phonological output when repeating words or nonwords, reading, or naming pictures. His errors in both languages were similar, except that he frequently metathesized root consonants when speaking or writing Arabic. Metatheses in French were quite rare, and he also never metathesized Arabic vowels or affix consonants. Thus, there is clear psycholinguistic evidence that Arabic consonantal roots are a distinct component of the Arabic mental lexicon.
Additional linguistic evidence for abstract consonantal roots in Arabic comes from consonant cooccurrence constraints that are restricted to the root consonants (McCarthy 1986). For example, forms with repeated final consonants (samam ‘to poison’) are well attested, while forms with repeated initial consonants (*sasam) are not found. McCarthy proposed that the OBLIGATORY CONTOUR PRINCIPLE (OCP) applies to Arabic root consonants such that repetition of adjacent consonants in the root is not allowed. In his analysis, words like samam are derived from underlyingly biconsonantal roots like /sm/. Repeated consonants appear on the surface as the result of multiple association proceeding from left to right, as in 2. In this way, a verb like *sasam can never be derived. Crucially, the OCP can refer only to verbal root consonants, as there is no prohibition of repeated consonants between roots and affixes.

\[ (2) \quad \begin{array}{cccc}
\text{s} & \text{m} \\
\text{C} & \text{V} & \text{C} & \text{V} & \text{C}
\end{array} \]

1.1. LEXICAL AND GRAMMATICAL INFLUENCES ON METALINGUISTIC WELL-FORMEDNESS JUDGMENTS. Previous studies of the influence of phonotactic constraints on native speaker well-formedness judgments have provided suggestive evidence for the psychological reality of abstract phonotactic constraints (e.g. Scholes 1968). In a study closely related to our own, Berent and Shimron (1997) asked Hebrew speakers to make preference judgments between nonsense words based on a pattern of unattested repeated consonants (e.g. sisem) and nonsense words based on a pattern without such repetition (e.g. pisem). Like Arabic, Hebrew allows repeated consonants only at the end of a root. Across several experiments, Berent and Shimron found a consistent preference for nonsense words without repetition of consonants, in support of the psychological reality of the segmental OCP. It is unclear, however, whether an abstract phonotactic constraint is necessary to explain these results. Since Hebrew contains no verb forms that start with sis but some words that start with pis, the judgments may have come from simple analogy to known lexical items.

There are only two studies that we know of that have directly compared the influence of grammatical factors and lexical factors in well-formedness judgments. These studies have concluded that phonotactic constraints are not necessary to explain well-formedness judgments. Ohala & Ohala 1986 compared the influences of phonotactic constraint violation and analogy on well-formedness judgments. The authors found that a measure of the similarity of their novel words to the words of English (taken from Greenberg & Jenkins 1964) provided a better account of the well-formedness judgments than a measure based on phonotactic constraint violations. Coleman and Pierrehumbert (1997) examined acceptability judgments for English nonwords containing a variety of grammatical and ungrammatical consonant clusters. They used a probabilistic parser to generate an expected probability for each nonword based on the frequency of the nonword’s onset and rime constituents. Illegal clusters were assigned a marginal low probability. Coleman and Pierrehumbert found acceptability judgments were more accurately predicted by the phonotactic probability metric than metrics based on constraint violation alone. However, in neither study was it clear that phonotactic constraint violation made no contribution to the judgments. Rather, lexical factors were shown to influence judgments, and lexical statistics and ungrammaticality were confounded so there was no evidence in favor of independent grammatical influences.

There is a growing literature in cognitive psychology on the concept of WORDLIKENESS, the degree to which a novel word is similar to existing words in the lexicon.
Wordlikeness influences nonword processing and metalinguistic judgments of well-formedness (see Frisch et al. 2000 for a recent review). Psychological measures of wordlikeness commonly use one of two approaches. Either wordlikeness is grounded in analogical comparison of whole words (as in the analysis of Ohala & Ohala 1986) or in wordlikeness stems from the cumulative statistical probability of subword units (as in Coleman & Pierrehumbert 1997). Bailey and Hahn (2000) present evidence from experiments using English nonwords that suggest that both statistical probability and analogy independently influence well-formedness judgments. There is also some evidence that both types of wordlikeness influence phonological processing (Vitevitch & Luce 1999). In general, however, psychological studies of wordlikeness have focused on possible nonwords, and phonotactically illegal nonwords are avoided. The two factors of analogy and statistical probability can potentially account for the rejection of illegal nonwords in metalinguistic tasks, as illegal nonwords are dissimilar from all existing words, and illegal nonwords contain a subword sequence of zero probability. In our experiment, we attempted to examine analogy to existing words, phonotactic probability, and phonotactic constraint violation independently to explore whether one or all of these factors influence well-formedness judgments.

1.2. GRADIENT CONSTRAINTS IN THE ARABIC ROOTS. Greenberg (1950) studied the phonotactics of the Arabic verbal roots by examining the statistical patterning of root consonant combinations in Lane’s 1863 dictionary. Greenberg showed that Arabic consonants divide into ‘sections’ of homorganic consonants that tend not to cooccur within the same root. McCarthy (1988, 1994) replicated Greenberg’s study, and characterized the coocurrence classes in terms of the combination of place-of-articulation features and the major manner feature [sonorant]. McCarthy (1988) accounts for the constraint against repeated homorganic consonants as an application of the OCP to the place feature tier, known as OCP-PLACE. Note that the case of the repeated final consonant, as in 2, is a potential violation of OCP-place. However, like the segmental OCP, the OCP-place constraint is proposed to apply to the underlying root, and not the surface form, and so there is no violation in these cases.

The major cooccurrence classes discussed by Greenberg and McCarthy are shown in 3. In their analyses, consonants in any one of these classes are claimed to cooccur freely with consonants from any other class, and within any class consonants tend not to cooccur, with two exceptions. First, the velars (3c) cannot cooccur with the uvular approximants {χ, ў}, though they can cooccur with the other gutturals. Second, among the coronal obstruents, there are far more roots containing one fricative and one stop than roots containing two fricatives or two stops. Note that in 3b the capital letters, {T, D, S, Z}, are used to indicate the uvularized coronal obstruents, also known as emphatics.

(3) MAJOR COOCCURRENCE CLASSES

a. Labials = {b, f, m}
b. Coronal Obstruents = {t, d, T, D, θ, s, z, S, Z, ў}
c. Velars = {k, g, q}
d. Gutturals = {χ, ў, h, ў, ў}
e. Coronal Sonorants = {l, r, n}

In dictionary studies of Arabic phonotactics, cooccurrence restrictions between consonants are discovered by comparing the number of roots that contain a particular consonant pair with the number that would be expected if consonants were to combine at random. Significant deviations from chance indicate consonant pairs that may be subject to a constraint. Pierrehumbert (1993) used the ratio of observed frequency to
expected frequency, O/E, as a simple metric for examining cooccurrence rate. If O/E is greater than or equal to 1, then cooccurrence is unrestricted. If O/E is less than 1, there may be a constraint.

Upon closer examination of the quantitative details of the cooccurrence constraint, Pierrehumbert claimed that OCP-place is a dissimilation constraint, where cooccurrence is quantitatively dependent on the perceived similarity of homorganic consonants. Consider the cooccurrence patterns of roots involving initial /t/, shown in Table 1. Combinations of root-initial /t/ with /s/ are infrequent, and perceptually /t/ and /s/ are very similar, as they are voiceless obstruents at the same place of articulation. Combinations of /t/ with /l/ are more common, and /t/ and /l/ are less similar than /t/ and /s/, as /l/ is palato-alveolar and not alveolar. Due to a contrast in the major manner feature [sonorant], /t/ and /n/ are quite dissimilar, though both are coronal, and cooccurrence is common. Finally, the frequent combination of /t/ and /k/ demonstrates that the constraint is crucially sensitive to homorganicity, as /t/ and /k/ are otherwise highly similar (see Frisch et al. 1997 for a formal account).

There are a number of reasons why the cooccurrence patterns discovered in a statistical dictionary analysis may not reflect synchronic grammatical knowledge. A dictionary may be an inaccurate characterization of a native speaker’s mental lexicon. A dictionary may contain borrowings, obscure forms, and no longer used forms that are not part of the productive grammatical knowledge of native speakers. In general, a dictionary is a better reflection of the standardized written language than of the spoken language. Even if a dictionary provides a reasonable approximation of the lexicon, the lexicon itself is an amalgam of diachronic processes, and the words it contains are, in part, an accident of history. In addition, the set of existing lexical items, while an open class, is relatively static and nonproductive in comparison to other aspects of morphology. It is, consequently, important to determine whether the well-formedness judgments of native speakers agree with the dictionary patterns for two reasons. First, converging evidence will validate the use of a dictionary as an appropriate corpus for examining phonotactic constraints that are potentially quantitative. Second, the abstract nature of the Arabic constraint provides a crucial testing ground in which to examine the psychological reality of phonotactic constraints.

2. Method

2.1. Material. We constructed 254 novel verbal roots, taking into account the probability of consonants and consonant combinations, as well as the distribution of existing roots, using the set of native and assimilated triliteral roots taken from Wehr’s dictionary of Standard Arabic (Cowan 1979). Our stimuli were conceptually divided into three sets, each designed to test a different facet of Arabic phonotactics and the nature of well-formedness judgments.

Stimulus set I consisted of 160 novel verbs. Equal numbers of verbs were constructed that contained either OCP-place violations or not, had either relatively high or low expected probability, and were similar to relatively more or fewer existing roots that
could act as analogical models. In this stimulus set there were 20 stimuli per three-way categorization of OCP-place violation, expected probability, and similarity to exist-

<table>
<thead>
<tr>
<th>CONSTRAINT VIOLATION</th>
<th>EXPECTED PROBABILITY</th>
<th>NEIGHBORHOOD DENSITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCP violation</td>
<td>High, 0.12–0.25 (M = 0.171)</td>
<td>Dense, 11–20 (M = 13.2)</td>
</tr>
<tr>
<td>No violation</td>
<td>Low, 0.06–0.12 (M = 0.068)</td>
<td>Sparse, 1–10 (M = 8.2)</td>
</tr>
</tbody>
</table>

Table 2. Three way categorization of stimuli.

ing roots. These groups are summarized in Table 2. Roots with OCP-place violations contained at least one pair of homorganic consonants selected from the major cooccurrence classes given in 3. The expected probability of a novel root is the probability of independent combination of the consonants given their frequency in the dictionary by (Frisch 2000). For example, the expected probability of the novel root */bts/ = P(bCC/) × P(CTC/) × P(CCs/), where C is any consonant. The similarity of a novel root to existing roots is measured using a simple metric based on single segment substitution, called the NEIGHBORHOOD DENSITY (Greenberg & Jenkins 1964, Luce & Pisoni 1998). Neighborhood density is taken to be the number of existing roots that share two of the three consonants in the novel root in the appropriate serial positions. For example, the neighborhood density of the novel root */bts/ = N(btC/) + N(Cts/) + N(bsC/) where C is any consonant. Obviously, violations of OCP-place are infrequent in the lexicon. Consequently all the novel roots were relatively low in expected probability and were similar to very few existing roots, whether they contained constraint violations or not.

Stimulus set II was designed to examine whether the well-formedness judgments of native speakers of Arabic would differentiate systematic gaps from accidental gaps, taking lexical statistics into account. Stimuli with violations of OCP-place were compared with stimuli containing accidental gaps that do not belong to a coherent natural class of nonoccurring consonant pairs. Stimulus set II was composed of 40 novel verbs, 20 containing a consonant pair that is an OCP-place violation and 20 with no violation, each with a single nonoccurring consonant pair. The stimuli were equated on expected probability and neighborhood density. All of the stimulus set II verbs had expected probability in the very low probability group and neighborhood density in the sparse group of Table 2. The novel verbs in stimulus set II were also balanced for bigram probability. Bigram probability is the frequency of occurrence of a consonant pair, rather than an individual consonant. For example, the novel verb root */thfl/ does not contain an OCP-place violation, but does contain an accidental gap. In Wehr’s dictionary, there are no roots of the form /thCfl/, four roots of the form /Chfl/, and three roots of the form /tCfl/, where C is any consonant. This nonviolation root was matched with the novel root */tsbfl/, containing an OCP-place violation. In Wehr’s dictionary, there are no roots of the form /tsCfl/, four roots of the form /Csbfl/, and three roots of the form /tCfl/, where C is any consonant.

Stimulus set III was designed to examine whether acceptability judgments of native speakers of Arabic would differentiate among different degrees of OCP-place violation as a function of similarity, as claimed by Pierrehumbert (1993). Stimulus set III was composed of 54 novel verbs. The stimuli were created in groups of three, with each group balanced with every other group for expected probability and density as in stimulus set I, and for bigram probability as in stimulus set II. Each stimulus in the group had a critical consonant pair in one position in the root, and all other consonant pairs
were not OCP-place violations. The critical consonant pairs were evenly distributed among the possible positions ($C_1$ and $C_2$, $C_2$ and $C_3$, and $C_1$ and $C_3$). The critical consonant pairs covered a range of similarity levels according to the metric proposed by Frisch, Broe, and Pierrehumbert (1997). At the extremes, there were six stimuli that did not contain any OCP-place violations (similarity 0) and nine stimuli that contained one pair of repeated identical consonants (similarity 1). Note that verbs with repeated identical consonants as $C_2$ and $C_3$ are frequently found in Arabic (Greenberg 1950). As discussed above, McCarthy (1986) analyzes these verbs as the result of left-to-right autosegmental spreading from an underlying biconsonantal root. These final geminate verbs are therefore not considered to contain OCP-place violations, as homorganic consonants are not repeated in the root. The remaining 39 stimuli contained a single OCP-place violation in the critical pair for different levels of similarity. This stimulus set thus provides a strong test of Frisch et al.’s (1997) account.

2.2. Participants. Thirty native speakers of Jordanian Arabic, living in Amman, were paid for their participation in this study. The participants ranged in age from 16 to 57. All participants were well educated members of the second author’s home community, and are speakers of the Ammani-Jordanian dialect. There are no aspects of the dialect that differ from Standard Arabic sufficiently to introduce major differences between the pattern of OCP-place violations in the dictionary and OCP-place violations in the dialect (see Zawaydeh 1999). The participants were also speakers of English and French with varying levels of fluency. While the diversity of this population may introduce additional undesirable variability into the data, any patterns that emerge despite such variability are likely to be robust reflections of the knowledge of Arabic possessed by all members of the community.

2.3. Procedure. The experiment was conducted entirely in Arabic. Instructions and stimuli were presented orthographically on a questionnaire in Arabic script. All verbs were given in the same $C_1aC_2aC_3a$ form, with diacritics for vowels included. The first page of the questionnaire provided instructions to judge each ‘made up verb’ using a 1–7 scale based on how much the verb sounded like a possible verb of Arabic. A rating of 1 was described as ‘Impossible. This word sounds terrible. It can’t be a verb of Arabic’. A rating of 3 was described as ‘Unlikely. This word sounds very strange. I doubt that it could be a verb of Arabic’. A rating of 5 was described as ‘Possibly. This word sounds a little strange, but it could be a verb of Arabic’. A rating of 7 was described as ‘Definitely. This word sounds very much like a verb of Arabic’. Participants were instructed to use 2, 4, and 6 for verbs that were between these suggested guidelines. Note that the wording of the questionnaire attempted to focus the participants on how the word sounded, and on the fact that it should be judged as a verb. In order to enhance the saliency of these words as verbal roots, the participants were encouraged to conjugate each verb while considering its wordlikeness. Each novel verb was presented on a single line of the questionnaire, along with the Arabic digits from 1 to 7. The questionnaire contained a randomized list of the stimuli that mixed together roots from the three stimulus sets. Filling out the questionnaire took about thirty minutes.

2.4. Analysis. We present an analysis of the data from 24 of the participants that showed gradient wordlikeness judgments. These 24 participants appeared to use a range of responses along the rating scale, as intended by the instructions. The other 6 participants appeared to use ‘categorical’ response strategies. Three of these participants rated nearly all the novel forms as unacceptable (rating 1). Three other participants used just
two ratings for all of the stimuli (1 and 3 for two participants and 3 and 5 for the third). Analysis of the theoretical significance of different response strategies in well-formedness judgments in Frisch & Zawaydeh 1997 and Frisch et al. 2001 have found that the categorical and gradient response strategies reflect the same underlying knowledge. Separate statistical tests for the other participant groups in our experiment will not be given here, but these tests support all of the conclusions presented below.

3. RESULTS

3.1. STIMULUS SET I: CONSTRAINT VIOLATION, PROBABILITY, AND BETWEEN WORD SIMILARITY. Three-way analysis of variance (ANOVA) over the three categories of OCP-place constraint violation, expected probability, and neighborhood density in stimulus set I revealed a significant main effect of OCP-place violation on mean ratings across subjects ($F_1(1,159) = 75.9, p < .001$) and across stimuli ($F_2(1,191) = 63.9, p < .001$). OCP-Place violations were rated about one point less acceptable on the scale than comparable nonviolations. There were no significant effects of variation in expected probability or similarity to existing roots as measured by neighborhood density. There were no significant interactions between any of the factors. Figure 1 shows the consistent effect of OCP-place violation across all expected probability and density categories. Note that, overall, both violations and nonviolations are generally rated on the lower half of the scale, suggesting the stimuli were not very wordlike to the participants. This may be due to the overall low expected probability and neighborhood density of the roots that was needed to orthogonally examine constraint violation and lexical factors.

Rating patterns for this stimulus set provide strong support for the psychological reality of the OCP-place constraint. OCP-place violation influenced the ratings, but the manipulation of expected probability and neighborhood density did not. However, there was considerable variation across the stimuli in the ratings given to stimuli with viola-
tions and those without violations. For example, it was not the case that all OCP-place violation stimuli were rated as less acceptable than all nonviolation stimuli. The presence or absence of an OCP-place violation predicts only about 30% of the variability in the mean responses ($R^2 = .29$), indicating there is a great deal of unexplained variation. Other factors not considered in this stimulus set may have influenced the ratings, and we return to this issue with two post-hoc analyses in §4. Further, the OCP-place violations generally contained a consonant pair that is unattested in the lexicon, while the nonviolations contained consonant pairs that were attested. This difference in bigram probability could be the explanation for the lower ratings for the violations. It was this possible account that stimulus set II was designed to eliminate.

3.2. Stimulus set II: accidental versus systematic gaps. Stimulus set II was designed for a simple ANOVA, with OCP-place violation as a factor. As in stimulus set I, there was an extremely strong effect of OCP-place violation. Novel roots with OCP-place violations were rated significantly less wordlike than roots without violations ($F_1(1,39) = 9.9, p < .01; F_2(1,47) = 15.1, p < .001$). Overall, the difference was about 0.8 on the rating scale. More importantly, we have clear evidence that Jordanian Arabic speakers do recognize systematic gaps that are violations of OCP-place as different from accidental gaps involving unrelated consonant pairs. Once again, there is a great deal of variability in responses, but OCP-place violation was the only clear influence on the ratings.

It is interesting to note that the difference between OCP-place violations and nonviolations in stimulus set II (0.8) is slightly smaller than the difference between OCP-place violations and nonviolations in equivalent stimuli in stimulus set I (1.0). Figure 2 shows mean ratings for OCP-place violations and nonviolations among equivalent stimuli in stimulus sets I and II (all stimuli in the very low probability and sparse

**Figure 2.** Comparison of mean ratings for systematic and accidental gaps (stimulus set II) with equivalent stimuli possibly containing no gap (stimulus set I). Error bars show 95% confidence interval for the mean.
density groups). Comparing the nonviolations between the two stimulus sets, mean ratings appear slightly lower for novel verbs in stimulus set II than in stimulus set I. The nonviolation verbs of stimulus set II contained accidental gaps, so they differed from the nonviolation verbs in stimulus set I that generally did not contain gaps. The difference in ratings is not quite statistically significant ($t(38) = 1.3, p = 0.1$) but the pattern is suggestive. Thus, the rating data we obtained may not be entirely devoid of influences from lexical statistics. The presence of an unattested consonant pair, even if it is an accidental gap, may explain some of the variation in ratings among the different stimuli.

3.3. Stimulus set III: Gradient OCP-place effects. Stimulus set III examined differences in well-formedness among OCP-place violations based on the gradient nature of the constraint itself (Pierrehumbert 1993; Frisch, Broe & Pierrehumbert 1997). The verbs in stimulus set III contained a variety of critical consonant pairs that covered a range of degrees of OCP-place violation. Mean ratings for each stimulus given the similarity of the critical consonant pair is shown in Figure 3. Nonviolations are indicated by filled circles at similarity 0. Stimuli with final geminates are indicated by filled boxes at similarity 1. Other stimuli with repeated identical consonants are indicated by x’s at similarity 1. Finally, stimuli with nonidentical OCP-place violations are indicated by open diamonds. A regression line for mean rating as a function of critical pair similarity is also shown. This regression is based only on the 39 nonidentical OCP-place violations (similarity strictly above 0 and below 1). The similarity based model of OCP-place is supported, as high similarity violations are rated less wordlike than low similarity violations. The effect of similarity on mean ratings for OCP-place violations in the regression model is statistically significant ($t(37) = 3.0, p < .01$).
Figure 3 shows a number of other interesting patterns, but it should be noted that these observations are based on only a few stimuli, and so must be considered tentative. First, it appears that novel verbs without OCP-place violations were rated slightly better than the verbs with the lowest similarity OCP-place violations. This is as predicted in the Frisch, Broe and Pierrehumbert (1997) account, where nonhomorganic consonants are predicted to be at the extreme lower end of the similarity scale. Also it appears that the three stimuli with final geminates were very highly rated, with higher mean ratings than any of the other stimuli. The novel words with the final geminate pattern (analogous to real words like samam) were recognized as containing a common pattern and judged highly acceptable despite the potential OCP violation of the geminate. This is contrary to Everett and Berent’s findings for similar stimuli in an experiment with Hebrew speakers (1997).

Finally, it appears that OCP-place violations involving identical consonants are not judged as bad as OCP-place violations involving highly similar but nonidentical consonants. This result is unexpected, as it is counter to the lexical pattern in the verbal roots, and the accounts of McCarthy 1994, Pierrehumbert 1993, and Frisch et al. 1997. Since there were very few stimuli in the experiment with identical consonant pairs, it is possible that some particular lexical factors specific to these stimuli are involved here and this is not a reliable result. It is worth noting, however, that there are dissimilation constraints analogous to OCP-place in other languages where identical consonant pairs are acceptable but highly similar and nonidentical consonant pairs are not. For example, the OCP-place constraint in Ngbaka permits repeated identical consonants while excluding homorganic consonant combinations that differ in just one distinctive feature (Broe 1995). The typology of laryngeal dissimilarity constraints explored in MacEachern 1999 is analogous. In some languages, both high similarity and identity are prohibited. In other languages, high similarity is prohibited, but identity is allowed. Clearly the special status of identity in dissimilation constraints is a fruitful topic for future phonological and psycholinguistic research.

Overall, stimulus set III provides strong support for the similarity-based account of OCP-place. For stimuli containing OCP-place violations, the similarity of the violating consonant pair did influence wordlikeness ratings. This provides support for an abstract but lexically grounded phonotactic constraint as a psychologically real part of the Arabic grammar. Further evidence for the abstractness of the constraint is found in the frequently attested samam pattern with a repeated final consonant. Novel verbs containing this pattern received very high wordlikeness ratings, higher than nonviolations with unrelated consonants, supporting the analysis of these verbs as containing underlyingly biliteral roots that do not violate the OCP (McCarthy 1986).

4. EXPLAINING THE REMAINING VARIABILITY WITH OTHER PREDICTORS. Well-formedness judgments for the three stimulus sets found robust effects of OCP-place violation, some evidence for the influence of lexical statistics like bigram probability, and no evidence that the similarity of the stimuli to the other roots in the lexicon affected judgments. These effects were the only ones included in the experimental design, so discovering further influences requires post-hoc analysis of the rating data. Post-hoc analysis of our experiment is motivated for two reasons. First, since there was a great deal of unexplained variation in the data, additional factors are likely at work that may influence the interpretation of the data given thus far. Second, if post-hoc analysis can discover additional influences, these can be investigated systematically or at least taken into account in future experiments on well-formedness judgments.
Given the recent increase in the use of psycholinguistic experimentation in phonological research, it is important to investigate potential confounding factors so they can be avoided.

A number of analyses were undertaken to consider a variety of potential influences; two analyses that produced results of theoretical interest are discussed in this section (see Frisch et al. 2001 for additional analysis and discussion). First, given the strong influence of OCP-place violations on the ratings, we performed an additional analysis that considered the effects of multiple constraint violations on native speaker judgments. Second, we attempted to investigate whether analogy to existing words influenced the judgments by taking into account the effect of particular salient lexical items on the ratings.

4.1. ANALYSIS I: MULTIPLE CONSTRAINT VIOLATIONS. It has been claimed that differences in degrees of acceptability between ungrammatical forms are to be expected, as some forms violate more constraints than others, and should therefore be considered less well-formed (Chomsky 1965, Everett & Berent 1997, Scholes 1968). Frisch (2000) contains an explicit model of multiple constraint violation in the case of the Arabic verbal roots. In this model, probabilistic cooccurrence constraints are multiplied to predict the distribution of roots with multiple constraint violations. The model’s predictions are borne out in the lexical patterns in the dictionary: Arabic verbs containing multiple OCP-place violations are found predictably less frequently than those containing a single violation.

In our experiment, there were some novel verbs in stimulus set I that contained multiple OCP-place constraint violations. Table 3 shows the number of stimuli, mean ratings, and standard deviation for stimuli containing zero, one, two, or three OCP-place violations. Overall, the mean rating appears to decrease as the number of violations increases, suggesting that multiple constraint violations are a psychologically real influence on well-formedness judgments. Admittedly, there are relatively few multiple violations, so once again the conclusion that can be drawn must be tentative. If the verbs with multiple violations are grouped together for statistical analysis, the difference in ratings between stimuli with one violation and stimuli with more than one violation just misses statistical significance ($t(143) = 1.3, p = .09$).

<table>
<thead>
<tr>
<th>NO. OF VIOLATIONS</th>
<th>N</th>
<th>MEAN</th>
<th>STD. DEV.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>109</td>
<td>3.56</td>
<td>±0.71</td>
</tr>
<tr>
<td>1</td>
<td>127</td>
<td>2.65</td>
<td>±0.78</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>2.48</td>
<td>±0.71</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>2.27</td>
<td>±0.56</td>
</tr>
</tbody>
</table>

Table 3. Effect of multiple violations of OCP-place on wordlikeness ratings.

4.2. ANALYSIS II: SIMILARITY TO SPECIFIC LEXICAL ITEMS. All of the stimuli in the experiment were similar to only a very few existing roots, so perhaps there was little opportunity for analogical effects on wordlikeness to be found. This is a relatively unstudied area in psycholinguistics, but at least one perceptual task has shown a sharp decrease in word similarity effects as the nonwords get further from actual lexical items (Wurm & Samuel 1997). The measure of lexical similarity we used, neighborhood density, determines the similarity of a stimulus nonword to the lexicon as a whole and has been shown to influence wordlikeness judgments for English (Greenberg & Jenkins 1964, Bailey & Hahn 2000). It is also conceivable that similarity to a specific salient lexical item, rather than similarity to the lexicon as a whole, would influence the well-formedness judgments of our participants. To explore this hypothesis, we tapped the
intuitions of the second author, who is a native speaker of the same dialect as the participants in our experiment. The second author read through the stimuli, and marked all of those that, when she read them, readily prompted a particular word to come to mind. In some cases, these highly similar words were not derived from the verb system, and thus were not contained in our dictionary of verbal roots. Despite the claim that OCP-place applies only to verbal roots, this fact suggests that future studies should use a broader lexicon when constructing Arabic stimuli in order to better control for lexical effects.

Mean ratings and the number of stimuli that were OCP-place violations or not, and were nonwords that evoked particular similar lexical items or not, are shown in Table 4. There is a very robust effect of the presence of a particularly salient lexical neighbor,

<table>
<thead>
<tr>
<th>NEIGHBOR/OCP GROUP</th>
<th>N</th>
<th>MEAN</th>
<th>STD. DEV.</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Salient Neighbor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No OCP Violation</td>
<td>82</td>
<td>3.42</td>
<td>± 0.66</td>
</tr>
<tr>
<td>OCP Violation</td>
<td>127</td>
<td>2.50</td>
<td>± 0.61</td>
</tr>
<tr>
<td>Salient Neighbor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No OCP Violation</td>
<td>27</td>
<td>4.00</td>
<td>± 0.69</td>
</tr>
<tr>
<td>OCP Violation</td>
<td>18</td>
<td>3.42</td>
<td>± 1.18</td>
</tr>
</tbody>
</table>

Table 4. Influence of salient existing word on wordlikeness ratings for novel roots.

especially for the stimuli containing OCP-place violations. The mean difference between stimuli with similar neighbors and those without is statistically significant for stimuli without OCP-place violations ($t(107) = 3.9, p < .001$) and stimuli with violations ($t(143) = 5.1, p < .001$). Thus, it appears that some of the variability in responses across our stimuli can be accounted for by the analogical influence of particularly salient similar lexical items on well-formedness judgments. Given the strong influence of a salient neighbor on well-formedness judgments, we repeated all of the previous analyses presented in this article, taking this additional factor into account. None of the original results were contradicted, and it appears that the influence of a salient neighbor on well-formedness judgments is independent of the variables that were intentionally manipulated in the experiment.

5. Discussion. The experiment and subsequent analyses provide strong support for the validity of dictionary studies of Arabic phonotactics. Native speaker judgments reflect detailed quantitative patterns of OCP-place constraint violation. Overall, there was a dispreference for roots containing repeated homorganic consonants. In other words, OCP-place is a psychologically real synchronic constraint. Stimulus set I showed Jordanian Arabic speakers are sensitive to OCP-place violations, independent of superficial lexical factors. Stimulus set II demonstrated that Arabic speakers differentiate accidental and systematic gaps, crucially ruling out any simple frequency based account of OCP-place effects on well-formedness judgments. Knowledge of the OCP-place constraint reflects knowledge of the abstract homorganic natural classes. Stimulus set III showed that native speaker judgments are influenced by the similarity of consonant pairs that are OCP-place violations, as proposed by Frisch et al. (1997). Post-hoc analysis further supports the psychological reality of an abstract OCP-place constraint by showing that multiple constraint violations further reduce the well-formedness of novel words for native speakers.

5.1. Well-formedness judgments and abstract grammatical constraints. The wordlikeness judgments we obtained demonstrate the psychological reality of the consonant cooccurrence constraints of Arabic. Judgments reflect the natural classes defined by homorganicity, but judgments are also influenced by the similarity of consonant
pairs that violate OCP-place. The effect of similarity on cooccurrence in the lexicon is gradient. Given our results, both categorical and systematic gradient influences appear to be psychologically real. The general lack of effect of neighborhood density suggests that the knowledge of acceptable and unacceptable phonotactic patterns is abstracted away from the individual roots in the lexicon, and so does not reflect straightforward analogy based on the comparison of a novel root to the set of lexical items. Thus, we simultaneously find support for feature-based linguistic categories (the natural classes of phonological theory) and distribution-based quantitative influences (gradient constraints). These results suggest that the Arabic consonant cooccurrence constraints are emergent generalizations derived from patterns in the lexicon.

In the post-hoc analysis, some influence of individual similar lexical items was found, and it is worthwhile to consider this influence more carefully. In this experiment, the metalinguistic judgments of native speakers were collected, with the goal of finding supporting evidence for one of the possible analyses of the phonotactic patterns of Arabic. Like any other linguistic behavior, well-formedness judgments are influenced by a variety of grammatical and nongrammatical factors (Schütze 1996). Crucially in this case, there were independent influences of OCP-place violation and the existence of a particular similar lexical item on the ratings. Since independent influences were found, it cannot be the case that knowledge of OCP-place violation reduces to similarity to particular lexical items. Intuitively, it is relatively straightforward to conclude that the influence of an OCP-place violation on the ratings is a grammatical factor, and that the influence of a particular similar lexical item on the ratings is a nongrammatical factor. For example, any psychology experiment involving judgments about category membership for stimuli, whether linguistic or nonlinguistic, will likely be influenced by similarity to particular category members (e.g. Allen & Brooks 1991). But demonstrating that there is a distinction between grammatical and nongrammatical factors does not negate the fact that both factors influenced well-formedness judgments and therefore that both factors are potentially linguistically relevant. There is no reason to expect that grammatical knowledge will be the only influence in loan word adaptation, novel word creation, diachronic language change (e.g. analogical leveling), or any other linguistic behavior. In our experiment, a great deal of care was taken to demonstrate that it is necessary to posit an abstract OCP-place constraint independent of nongrammatical influences. Similarly, other studies of language and linguistic behavior must be careful to consider the possibility that apparently grammatical factors can be reduced to nongrammatical ones.

5.2. IMPLICATIONS AND FUTURE DIRECTIONS. Native Arabic speakers’ judgments of wordlikeness for novel verbs provide important evidence that an abstract phonotactic grammar is a psychologically real component of linguistic competence. In the case at hand, experimental data was collected for a large number of speakers using carefully controlled stimuli. This data shows that knowledge of phonotactic constraints is independent from simple lexical statistics. We also found, however, that phonotactic knowledge is more sophisticated than a symbolic description of possible words, as gradient patterns are also encoded. Consequently, the traditional, example driven methodology of linguistic analysis can discover only the grossest phonotactic patterns that native speakers have learned. Quantitative generalizations can be uncovered only if a substantial corpus, such as a dictionary, is examined. We have provided some evidence that a dictionary can be used as a reasonable approximation to the lexical knowledge possessed by a typical native speaker.

We have argued that the consonant cooccurrence constraints of Arabic are emergent generalizations over the lexicon. In an emergent grammar, phonotactic knowledge is
abstract, but not too abstract. The emergent constraints, based on lexical patterns, are language specific and learned. While these quantitative constraints clearly cannot be universal in their full quantitative detail, they are not incompatible with the presence of universals in phonotactics or elsewhere in the grammar. It is beyond the scope of this article to treat the important issues surrounding the integration of universal and language specific constraints, quantitative and categorical constraints, and acquisition guided by innate linguistic predispositions versus the discovery of language-specific patterns. It is apparent that native speakers are aware of quantitative constraints in the lexicon that are distinct from other nongrammatical influences. Theories of generative grammar are traditionally nonquantitative (with rare exceptions, e.g. Anttila 1997, Boersma 1998, Frisch et al. 1997). Since the cooccurrence constraints in Arabic apply to well-defined natural classes whether they are categorical or statistical tendencies, significant generalizations about Arabic phonotactics will be left unexplained unless quantitative constraints are allowed (see Abney 1997, for related arguments). The existence of psychologically real emergent phonotactic constraints demonstrates that knowledge of linguistic patterns goes beyond a set of universal principles that are cognitively instantiated by a set of parametric alternatives. For phonotactics, at least, the lexicon provides a rich enough source of data for phonological generalizations to be derived directly from lexical patterns.

REFERENCES


Prunet, Jean-François; René Beland; and Ali Idrissi. 1999. The mental representation of Semitic words. Montreal: University of Quebec, MS.


