How Many Grammars Am I Holding Up? Discovering Phonological Differences between Word Classes

Adam Albright
Massachusetts Institute of Technology

1. Introduction

It is a well established fact that languages may exhibit phonological differences between different classes of words. For example, nouns and verbs may show systematic differences with respect to stress placement (English: Chomsky & Halle 1968, chap. 3; Lenakel: Hammond 1984), accentual or tonal patterns (Fukuoka Japanese: Smith 1999, 2001; Hausa: Newman 2000), prosodic root shapes (Classical Arabic: McCarthy 2005), and even the distribution of segmental contrasts (English: Sereno & Jongman 1990; Kelly 1992; Sereno 1994). Within a single part of speech, words belonging to different subclasses may also have distinct phonological properties. For example, it has been observed that English verbs that allow dative shift and verbs that can occur in combination with particles both show a strong tendency to be monosyllabic (Grimshaw & Prince 1986; Jackendoff 1997:542), while in Turkish, place names and proper nouns show a unique stress pattern not found in common nouns (Sezer 1981; Inkelas *et al.* 1997). Different inflectional classes may also be associated with distinguishing phonological properties, either uniquely or as a statistical tendency. To take just a few examples, phonological differences between regular and irregular classes have been widely discussed for English verbs (Bybee & Moder 1983; Albright & Hayes 2003), German nouns (Köpcke 1993), and Italian verbs (Davis & Napoli 1994; Albright 2002; Boyé & Montermini 2007). Finally, some languages show systematic phonological differences between apparently arbitrary sets of words that cannot be distinguished by any clear syntactic or morphological criteria; such languages include Japanese (Itô & Mester 1995), Turkish (Itô *et al.* 1996; Inkelas *et al.* 1997), Assamese and Yine (Pater, in press).

In some cases, phonological differences can be fruitfully analyzed as a consequence of the contexts that different classes of words occur in, thus avoiding the need to stipulate class-by-class differences directly (McCarthy 2005). However, in many cases the difference cannot be attributed to the influence of surrounding morphological contexts, but rather appears to be synchronically arbitrary. To handle such cases, it is standardly assumed that phonological rules or constraints can be relativized to particular sets of words (Chomsky & Halle 1968; Kenstowicz & Kisseberth 1977:77–83; Itô & Mester 1995; Smith 1999; Pater, in press; etc.). In rule-based phonology, the unit of relativization is nearly always the rule: through use of indices a process can be restricted to an individual class, or a particular class can be exempted from undergoing the process. Within Optimality Theory (Prince and Smolensky 1993/2004) a broader range of approaches have been pursued, with some authors advocating class-specific constraints intermixed with general ones (Itô & Mester 1995), and others advocating entirely separate subgrammars or co-phonologies (Inkelas *et al.* 1997; Anttila 2002).

All of these approaches face a common challenge, however, which is to explain how learners determine that relativized grammars or constraints are needed in the first place. Two popular strategies seem promising. On the one hand, one could adopt a greedy top-down approach, assuming that all patterns are fully general and subdividing the grammar only on the basis of overt evidence. Such evidence could consist of statistical distributional differences or different alternations (Inkelas *et al.* 1997). Under this first approach, learners would posit word class differences cautiously and reluctantly, avoiding needless proliferation of relativized rules/constraints/grammars. Conversely, one could adopt a bottom-up approach, assuming at the outset that each class is unique and unifying them only as

Thanks to Michael Becker, Edward Flemming, Morris Halle, Bruce Hayes, Michael Kenstowicz, Stefanie Shattuck-Hufnagel and Donca Steriade for helpful discussion on various portions of the material presented here; all errors and misunderstandings are, of course, my own. Thanks also to Corinna Hui and Sophie Wong for help preparing materials and collecting data in the English experiments reported in section 3.

© 2008 Adam Albright. Proceedings of the 26th West Coast Conference on Formal Linguistics, ed. Charles B. Chang and Hannah J. Haynie, 1-20. Somerville, MA: Cascadilla Proceedings Project.

appropriate. Under this approach, learners would start with rich and detailed representations, and a conservative inductive learning procedure would notice whatever cross-cutting regularities are present in the input. For a bottom-up approach, word class differences could be seen as an automatic and necessary by-product of learning.

In this paper, I will argue that neither a fully top-down nor bottom-up approach is quite right. I will contrast two cases that appear to show puzzlingly contradictory effects. The first example comes from Spanish, where it appears that speakers strictly limit generalizations about vowel alternations by conjugation class, even in cases where the evidence for class-by-class differences is weak or non-existent. The second example concerns segmental phonotactic differences between nouns and verbs in English. For this case, I will present evidence that speakers ignore certain statistically significant differences between classes. In both cases, the challenge is to explain why speakers seem to have selected a grammar that is in some sense inappropriately restricted, given the data of the language. I will argue that this paradox can be resolved by assuming that learners decide whether or not to posit class-by-class differences rather coarsely for the language "as a whole", rather than for individual processes or inflectional classes. As we will see, a consequence of this coarse level of decision is that learners may end up forced to make morphological restrictions in parts of the lexicon that do not demand them, or may fail to posit morphological restrictions even though they would be useful in a particular corner of the language. I claim that this preference for consistency of analysis across the entire language can be captured by a hierarchical model of grammar inference.

2. Spanish conjugation classes

Most varieties of Spanish have a standard inventory of five phonemic vowels (/i/, /e/, /a/, /o/, /u/); also of relevance here will be the two diphthongs /je/ and /we/, which are disfavored in stressless syllables. Traditional grammars distinguish three major conjugation classes of Spanish verbs, which are characterized by the theme vowel that occurs between the stem and the present tense indicative person and number endings. These are illustrated in (1), with the theme vowel highlighted in bold face.

(1) Spanish conjugation classes

a. *habl**á**r* 'speak' (Class 1)

1sg	*hábl-o*	1pl	*habl-**á**mos*
2sg	*hábl-**a**s*	2pl	*habl-**á**is*
3sg	*hábl-**a***	3pl	*hábl-**a**n*

b. *com**é**r* 'eat' (Class 2)

1sg	*cóm-o*	1pl	*com-**é**mos*
2sg	*cóm-**e**s*	2pl	*com-**é**is*
3sg	*cóm-**e***	3pl	*cóm-**e**n*

c. *viv**í**r* 'live' (Class 3)

1sg	*vív-o*	1pl	*viv-**í**mos*
2sg	*vív-**e**s*	2pl	*viv-**í**s*
3sg	*vív-**e***	3pl	*vív-**e**n*

In theory, one might expect words with all five vowels to occur freely in each of the three conjugation classes (C_0iC_0-ar, C_0eC_0-ar, C_0aC_0-ar, ..., C_0iC_0-er, C_0eC_0-er, ..., C_0iC_0-ir, C_0eC_0-ir, C_0aC_0-ir, etc.). In point of fact, there are rather extreme differences between the conjugation classes with respect to the distribution of vowels in the final syllable of the stem (Harris 1969, chap. 4; Boyé and Hofherr 2006). This is illustrated in Figure 1 with counts from LEXESP (Sebastián *et al.* 2000), including all verbs with a frequency of at least 10 in the corpus. The counts reveal that high, mid and low vowels occur with roughly equal frequency in class 1, with roughly 350-450 of each (recall that there is only one low vowel, but two each of mid and high vowels). In classes 2 and 3 on the other hand, low vowels are

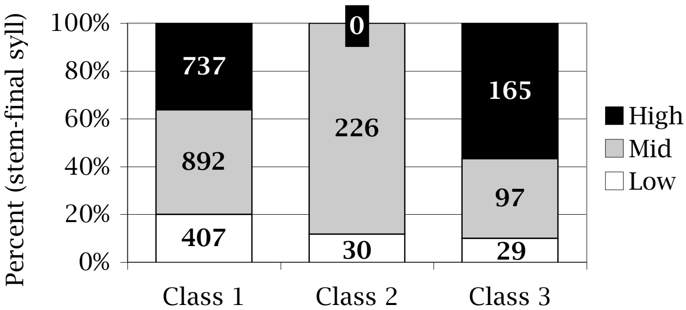

Figure 1: Distribution of stem-final vowels across conjugation classes (LEXESP counts)

statistically underrepresented. Even more striking, class 2 completely prohibits high vowel verbs, while class 3 strongly favors them.[1]

There are several sources of evidence that speakers are aware of these distributional differences, particularly with regard to the mid/high contrast. Historically, many class three verbs with mid vowels (and especially /o/) have been eliminated, either by raising to avoid the dispreferred combination (e.g., Latin *mollīre* ⇒ *mullir* 'soften'; Penny 2002:160), by changing conjugation class (older class 3 *colorir* 'color' ⇒ class 1 *colorear*) or by simply falling out of use as inflected verbs (e.g., *despavorir* 'be frightened', now confined mainly to the participial form *despavorido* 'terrified'). Furthermore, as we will see below, the handful of mid verbs that do remain in this class almost all undergo synchronic alternations (*e*/*o* ∼ *jé*/*wé*, or *e* ∼ *í*). These facts suggest a grammatical sensitivity to conjugation class, in which vowel contrasts are relativized to different classes either via different constraints or different grammars. This is illustrated, simplifying somewhat to assume a categorical distribution, in (2a)–(2b).

(2) Grammatical sensitivity to conjugation class

a. Relativized constraints in a single grammar

$$\begin{array}{c} \text{*[+high]/Class 2, *[+low]/Class 2, *[+low]/Class 3, \ldots} \\ | \\ \mathscr{F} \\ | \\ \text{*[+high]/Class 1, *[+low]/Class 1, \ldots} \end{array}$$

b. Distinct grammars for different conjugation classes

- Class 1: $\mathscr{F} \gg$ *[+high], *[−high, −low], *[+low]
- Class 2: *[+high] $\gg$ *[+low] $\gg \mathscr{F} \gg$ *[−high, −low]
- Class 3: *[+low], *[−high, −low] $\gg \mathscr{F} \gg$ *[+high]

The conjugation classes also differ substantially in the rate at which their members participate in vowel alternations. Spanish verb paradigms exhibit several well-known vowel alternations between unstressed mid vowels and stressed diphthongs (diphthongization: (3a)) or between mid and high vowels (raising: (3b)). In both cases, the alternation is triggered when stress falls onto the root by the regular principles of Spanish stress placement (Harris 1969, 1992, 1995; Eddington 2000; Oltra Massuet & Arregi 2005). The choice of diphthongization vs. raising is not predictable, and the existence of many non-alternating verbs ((3c)) shows that neither alternation is phonotactically necessary (i.e., stressed mid vowels are frequently tolerated, even in verbs). These alternations have received numerous treatments in the literature (Harris 1969, 1977, 1985; Brame & Bordelois 1973; Hooper 1976; Schuldberg 1984; Garcia-Bellido 1986; Carreira 1991; Eddington 1998), though the exact mechanism used to derive them is not our primary concern here.

[1] This is perhaps related to the fact that the theme vowel for class 2 is mid, and for class 3 is high.

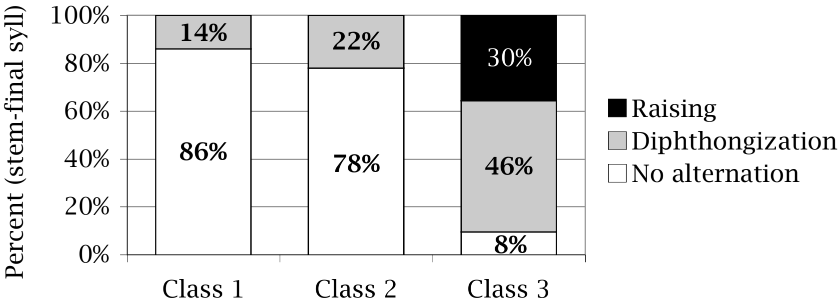

Figure 2: Distribution of mid-vowel alternations across conjugation classes (LEXESP counts)

(3) Alternations conditioned by stress (lexically restricted)

a. Diphthongization of *e*, *o*

i. *sentar* 'seat'	
s[**jé**]*nt-o*	*s*[**e**]*nt-ámos*
s[**jé**]*nt-as*	*s*[**e**]*nt-áis*
s[**jé**]*nt-a*	*s*[**jé**]*nt-an*

ii. *contar* 'count'	
c[**wé**]*nt-o*	*c*[**o**]*nt-ámos*
c[**wé**]*nt-as*	*c*[**o**]*nt-áis*
c[**wé**]*nt-a*	*c*[**wé**]*nt-an*

b. Raising of *e*

i. *pedir* 'request'	
p[**í**]*d-o*	*p*[**e**]*d-ímos*
p[**í**]*d-es*	*p*[**e**]*d-ís*
p[**í**]*d-e*	*p*[**í**]*d-en*

c. Neither

i. *rentar* 'rent'	
r[**é**]*nt-o*	*r*[**e**]*nt-ámos*
r[**é**]*nt-as*	*r*[**e**]*nt-áis*
r[**é**]*nt-a*	*r*[**é**]*nt-an*

ii. *montar* 'mount'	
m[**ó**]*nt-o*	*m*[**o**]*nt-ámos*
m[**ó**]*nt-as*	*m*[**o**]*nt-áis*
m[**ó**]*nt-a*	*m*[**ó**]*nt-an*

As with the distribution of height contrasts, the distribution of mid vowel alternations differs considerably across the three conjugation classes, as shown in Figure 2. Corpus counts from LEXESP reveal that diphthongization is a minority pattern in classes 1 and 2 but robust in class 3, and that raising is confined exclusively to class 3. Furthermore, nearly every verb in class 3 alternates in one way or another (Harris 1969:106–108).

As above, there are indications that speakers are sensitive to class-by-class differences in mid vowel alternations. Historically, many class 1 and 2 verbs have lost diphthong alternations, almost always generalizing the mid vowel (Penny 2002:157; Morris 2005): class 1 *priesta* ⇒ *presta* 'lend-3sg', class 2 *suerbe* ⇒ *sorbe* 'sip-3sg'.[2] In class 3, on the other hand, there has been no tendency for regularization, which reflects (and is reflected by) the relative lack of non-alternating stems in this class.

Inflection class differences in rate of diphthongization can also be observed synchronically in speakers' willingness to extend alternations to novel ("wug"; Berko 1958) words. In general, diphthongization is unproductive in class 1, and speakers are reluctant to extend alternations to novel verbs (Bybee & Pardo 1981). However, diphthongization is moderately robust in certain phonological contexts ("islands of reliability"; Albright 2002). For example, class 1 verb stems ending in *-rr* are statistically very likely to diphthongize, and speakers are correspondingly more willing to apply diphthongization to novel verbs like *lerrar* (Albright *et al.* 2001).

Unfortunately, since classes 2 and 3 are themselves unproductive (novel verbs are assigned almost exclusively to class 1), it is not plausible to present speakers with verb stems in all three classes in order

[2] A systematic exception to the usual direction of change involves derived verbs that are related to nouns or adjectives with diphthongs. These tend to retain the diphthong: *m*[wé]*bles* 'furniture' → *am*[we]*blár* 'furnish'.

to test directly for differences in the rate of diphthongization across the three classes. However, Albright et al. (2001) present at least some indirect evidence that generalization is sensitive to conjugation class. They attempt to model the likelihood of diphthongization of novel class 1 verbs using two different training sets: one including verbs from class 1 only, and one including verbs from all three classes. They find that on the whole, the choice of training set does not alter the results significantly. However, if we focus on responses for non-alternating (mid vowel) outcomes, a qualitative difference does emerge: the model based on all three classes does not correlate significantly with subjects' responses ($r(31) = .291$, $p = .1$), while a model based on class 1 alone does achieve a significant correlation ($r(31) = .385$, $p < .05$). Thus, it appears that the best model of Spanish speakers' knowledge of diphthongization is one in which alternations are learned separately for each conjugation class.

A more striking demonstration of class-by-class differences can be seen in cases where generalization is "unjustifiably" limited to a particular class. Recall (Figure 2 above) that unlike in classes 1 and 2, mid-vowel verbs in class 3 nearly all alternate somehow, either by raising or by diphthongization. As it turns out, class 3 differs in yet another salient way from the other two classes: it has practically no verb stems with the back mid vowel /o/. This can be seen in the table in (4), which shows LEXESP corpus counts of mid vowel verbs across the three conjugation classes.

(4) Distribution of front and back mid vowels across the conjugation classes

	Class 1	Class 2	Class 3
/e/	1096	229	124
/o/	677	90	8

A more accurate depiction of class 3, then, would be to say that in this class front vowel (/e/) verbs are well attested, and nearly all of them alternate. By contrast, speakers have very little evidence about the behavior of back mid vowel verbs in this class. There are two high-frequency verbs that diphthongize (*dormir* ∼ *duermo* 'sleep-inf/1sg', *morir* ∼ *muero* 'die-inf/1sg'), and one high-frequency verb with irregular 1sg. -*g*- but no diphthongization (*oír* ∼ *oigo*, *oye* 'hear-inf/1sg/3sg'). The remaining five class 3 /o/ verbs in the corpus are currently undergoing a change of raising to /u/ (older *podrir* ⇒ newer *pudrir* 'to stink, rot'), have switched conjugation class (older class 3 *colorir* ⇒ newer class 1 *colorear* 'to color'), or occur mainly in infinitive and participial forms (*abolir* 'to abolish', *despavorir* 'to be terrified', *descolorir* 'to de-color'). Therefore, even the count of 8 class three /o/ verbs overestimates the availability of data about such verbs.

Based on the evidence of essentially just three existing verbs, speakers should in principle be free to assume that class 3 /o/ verbs behave exactly like class 1–2 /o/ verbs—that is, that by default they do not alternate. This would correspond to a learning strategy in which learners assume that all processes are fully general, positing class-by-class differences only on the basis of strong positive evidence that the subdivision is well-motivated. However, there is evidence that speakers do not generalize from classes 1–2 to class 3. As noted in the preceding paragraph, there are several class 3 verbs that are not typically used in present tense inflected forms (*abolir*, *despavorir*, *descolorir*). If the need arose to produce inflected forms of these verbs, speakers would very likely need to consult their grammars, since they have no direct (lexicalized) knowledge of how the vowels of these particular verbs behave under stress. Thus, such verbs provide an opportunity for a real-world "wug test" involving unattested forms of existing words.

Albright (2003) presented a variety of existing mid-vowel verbs to adult native Spanish speakers, in order to elicit stressed inflected forms (the 3pl or 1sg). The responses for verbs like *abolir* showed massive uncertainty: "*abuel... abuelen... abo... ellos abuelen? abolen...* ?" This uncertainty parallels a claim often found in descriptive sources that the relevant forms of these verbs simply do not exist (e.g., Butt 1997). If the grammar of vowel alternations did not depend on conjugation class, there would be no problem producing these forms, since class 1 verbs show a default pattern of non-alternation with no uncertainty, and class 3 provides no counterevidence to prevent this pattern from being generalized. Instead, it appears that speakers have no default rule for class 3 /o/ verbs, and are forced to decide based the scant (and contradictory) evidence of *dormir* 'sleep', *morir* 'die', and *oír* 'hear'. These three examples are evidently not enough to support confident rule inference, and the fact that speakers cannot look beyond them to adduce support for non-alternating outputs (*abolen*, *despavoren*, *descoloren*) provides further evidence for the claim that the grammar of Spanish vowel alternations is subdivided by conjugation class. Since in this case there is no positive evidence that class 3 /o/ verbs behave differently from class 1 and 2 /o/ verbs, we are forced to conclude that speakers do not always require overt evidence to posit "split" grammars.

The data in this section all converge on a model of Spanish vowel phonology that is divided by conjugation class—not only for properties that do vary substantially across the classes (vowel height contrasts, diphthongization and raising of /e/) but also for properties that are not observably different across classes (diphthongization of /o/). One possible interpretation is that speakers are simply inherently biased or limited to form local, detailed generalizations that do not abstract over multiple word classes. A different interpretation, which will be defended here, is that the subdivision of Spanish phonology by conjugation class is a "global" decision: the fact that *many* properties are observably different across the classes motivates a division in which *all* properties are encoded separately. The evidence for this view comes from a case that shows the converse effect: when class-by-class differences are found for just a small number of properties, speakers evidently do not splinter their grammars in order to capture them.

3. English nouns and verbs

3.1. Background

Numerous studies have documented phonological differences between nouns and verbs in English (Kelly 1992; Sereno 1994; Berg 2000). The most frequently discussed difference concerns default stress placement (Chomsky & Halle 1968; Liberman & Prince 1977; Burzio 1994; Hayes 1995). In general, verbs receive stress on their final syllable if it contains either a complex nucleus or a complex coda (*obéy*, *predíct*), whereas nouns receive final stress only if the nucleus is complex (*arráy*, but *édict*)—and not even always then (*éxploit* (n.), **exploít*). Final syllables with simplex nuclei and single codas are normally unstressed regardless of part of speech (*édit*, *rábbit*). Because of the different treatment of coda clusters, words ending in two consonants may receive final stress as verbs but penultimate stress as nouns, yielding minimal pairs such as *recórd* (verb) vs. *récord* (noun). Experimental work using a variety of paradigms has shown that speaker behavior reflects this difference, and novel words with final stress are preferentially interpreted as verbs (Cassidy & Kelly 1991; Guion *et al.* 2003).

A variety of segmental differences also occur between nouns and verbs. For example, verbs are somewhat more likely to end in final voiced fricatives than nouns are, for the historical reason that verbs had a vowel-initial infinitive suffix that triggered intervocalic voicing of fricatives, while many nouns lacked an overt suffix in the singular (Jespersen 1942:§§12.1–12.4). This has created a number of noun/verb pairs that differ in fricative voicing, of varying degrees of synchronic analyzability (*sa*[f]*e* (n./adj.) vs. *sa*[v]*e* (v.); *réfu*[s]*e* (n.) vs. *refú*[z]*e* (v.); *grass* (n.) vs. *graze* (v.)). Fricative voicing is applied somewhat productively in derived verbs,[3] and deverbal nouns are also sometimes subjected to devoicing.[4] Verbs may also be somewhat more likely to have stressed front vowels than nouns are (Sereno & Jongman 1990; Berg 2000). Berg (2000) performed counts of stressed vowels among the 1000 most frequent monomorphemic nouns and verbs in CELEX (Baayen *et al.* 1993), finding that nouns have a slight statistical preference for back vowels (53.4%) while verbs show a converse preference for front vowels (54.3%). This difference may in part reflect a historical umlaut process that fronted vowels with certain verb suffixes (e.g., *tooth* [tuːθ] (n.) vs. *teethe* [tiːð] (v.); *lock* [lak] (n.) vs. *latch* [læt͡ʃ] (v.)), and may also be partly coincidental. For a more in-depth discussion of noun/verb differences in English and evidence that speakers notice such differences see Kelly (1992).

All of the differences described here are statistical trends rather than absolute and categorical distinctions. Nevertheless, there is ample reason to believe that speakers can in principle notice and encode such tendencies. For example, the Spanish data discussed in the previous section, in which diphthongization is applied variably depending on the consonantal context, requires that speakers track the gradient probability of diphthongization in different environments. More relevant to English

[3]E.g., Middle English innovations *mou*[ð]*e*, *shea*[ð]*e*, *gla*[z]*e* (Jespersen 1942:202–203). Voicing still seems somewhat productive in present day English—e.g., *rebir*[ð]*ing* (Law and Order SVU, Episode 169, Nov. 21, 2006), related to the noun *bir*[θ], *bir*[θ]*s*; more on this below.

[4]E.g., Middle/Modern English innovations *belief*, *proof* (Jespersen 1942:200–201); Sapir (1921:75) cites a pronunciation of the noun *rise* as [raɪs] ("the ri[s]e of democracy").

phonotactics, numerous authors have demonstrated that speakers are able to judge gradient, intermediate degrees of acceptability of novel words according to the probability of the sequences that they contain (Scholes 1966; Coleman and Pierrehumbert 1997; Frisch, Large and Pisoni 2000; Bailey and Hahn 2001; Hayes and Wilson, in press), and that such differences also play a role in on-line processing (Kelly 1992; Farmer *et al.* 2006). This suggests the possibility that, as with Spanish vowels, knowledge of English phonotactics may be subdivided into noun and verb-specific grammars.

3.2. Experimental data

In order to test for gradient probabilistic knowledge of segmental differences between different parts of speech, two experiments were carried out gathering acceptability ratings of nonce words presented either as nouns or as verbs. For both experiments, batches of monosyllabic non-words were selected to embody a wide range of phonological structures and degrees of phonological well-formedness. Novel words were presented auditorily in a simple frame sentence, either as nouns ("[blɪg]. *This is a* [blɪg].") or as verbs ("[blɪg]. *I like to* [blɪg]."). Part of speech was counterbalanced across subjects, so that each subject heard half the items as nouns and half as verbs. After hearing the carrier sentence, subjects repeated the word aloud and then rated it as a possible word of English on a scale from 1 (= implausible as a word of English) to 7 (= would make a fine word of English). Subjects' repetitions were recorded and transcribed by two phonetically trained listeners. If the transcribers did not agree that the word had been perceived/repeated as intended, the rating from that trial was discarded. In the first experiment 170 items were presented for rating, and there were 205 items in the second experiment.

Overall, participants' ratings of noun presentations were highly correlated with ratings of verb presentations (Experiment 1 $r = .824$; Experiment 2 $r = .864$), indicating that varying the part of speech had relatively little impact on responses. This is not surprising, since none of the observed phonological differences between nouns and verbs are categorical, and indeed in some cases, they are quite small (recall the ≈7% difference for vowel backness in the Berg corpus counts cited above). Furthermore, the features that have been observed to differ across parts of speech (final fricative voicing, stressed vowel backness) are few compared to the set of properties that are common to all parts of speech in English—e.g., a dispreference for *#dw* onsets, the rarity of [θ]-final clusters (*hearth*, *wealth*, *month*), and so on. Given that the effect is likely to be small and centered on a particular subset of test items, it makes sense to focus the analysis on just those novel words that are actually expected to show the greatest noun/verb differences. In order to do this, we need a model that makes quantitative predictions about the relative well-formedness of particular sequences as nouns or as verbs.

3.3. A model of gradient well-formedness

One of the simplest classes of statistical models to track the the rate of attestation of particular sequences are *n*-gram models, which estimate the probability of a novel string based on the joint probabilities of substrings of length *n*. For example, a bigram model of the probability of the novel word [blɪk] would consider the probability of an initial [b], the probability of [l] after [b], the probability of [ɪ] after [l], the probability of [k] after [ɪ], and the probability of stopping after [k]. The probabilities of transitions from one segment to the next are multiplied to yield a combined estimate of the probability that they would co-occur in the same word (=JOINT TRANSITIONAL BIGRAM PROBABILITY). Studies have found that novel words with higher bigram probabilities (calculated in this or related ways) are easier to repeat quickly (Vitevitch *et al.* 1997; Vitevitch & Luce 1998, 2005), easier to evaluate as same/different (Vitevitch & Luce 1999), and are judged to be more wordlike (Vitevitch *et al.* 1997; Bailey & Hahn 2001).

Bigram models stated over literal segments ([b], [l], [ɪ], [k]) are insufficient for purposes of evaluating a wide variety of nonce words, since even a single unattested combination of sounds is enough to yield a probability of 0—yet speakers clearly prefer some unattested combinations over others (e.g., *#bn* > *#lb*; Berent et al. 2007). Albright (in prep.) proposes to overcome this problem by allowing the model to refer to natural classes of segments stated in terms of phonological features. Bigrams of adjacent segments are parsed into combinations of natural classes: [bl] = [voiced labial stop][lateral], [voiced labial stop][sonorant], [voiced stop][sonorant], [voiced][voiced], and so on. Typically there are many different ways to characterize a particular pair of segments in terms of natural classes, but not all of these combinations of classes are equally probable in the training data. For example, voiced stops are very often followed by sonorants in English ([ba], [bl], [br] common, [bd], [bt] rarer), but the broader class of voiced segments is not so often followed by another voiced segment, since combinations like [ak], [lt], etc., are also widely attested. Therefore it is necessary to decide, given a particular pair of segments, what the most likely characterization is as a combination of natural classes.

One strategy for picking the appropriate natural classes would be to simply take the most probable among the relevant combinations of classes—in this case, favoring [voiced stop][sonorant]. However, it turns out that this tends to overestimate the goodness of unattested combinations by letting well-attested combinations "shoe in" similar combinations with rare members. For example, [s] is often followed by coronal consonants in English (*st*, *sn*, *sl*), making the bigram [*s*][coronal] very probable. However, this makes the potentially fatal prediction that other [*s*][coronal] combinations such as [sr], [sz], and [sd] should also be relatively acceptable. A solution to this problem is to take into account the fact that choosing a particular combination of segments involves two components: we must choose the relevant combination of natural classes, and we must also instantiate the natural classes with the segments in question. In other words, the goodness of [st] depends not only on the probability of coronal consonants after [s], but also on the probability of choosing [t] as the particular coronal consonant. This has the effect of penalizing sweeping generalizations, since it is advantageous to choose natural classes that are as narrow as possible (i.e., that have their probability mass distributed over as few segments as possible). The definition of the probability of a pair of segments, as a function of both the probability of the natural classes and also the probability of choosing the particular segments, is given in (5). The combination of these scores to yield a probability for an entire word is given in (6).[5]

(5) Probability of a particular pair of segments *xy*

$$= \textit{argmax}\ \mathrm{P}([\text{class1 containing } x][\text{class2 containing } y]) \times \mathrm{P}(x \text{ among class1}) \times \mathrm{P}(y \text{ among class2})$$

(6) Probability of novel word [blɪk] = $\prod$ P(pairs in [blɪk])

As a preliminary test of the model's ability to capture gradient preferences for some sequences over others, we can examine its performance on a batch of 92 novel words used as a pretest in a study of past tense formation (Albright & Hayes 2003). As in the current study, subjects in that study rated items on a scale from 1 (implausible as an English word) to 7 (would make a fine English word); however, the part of speech was held constant, with all words being presented as verbs. The model was trained on all of the lemmas in CELEX with a token frequency greater than 0. As can be seen from Figure 3, the model achieves a reasonably good fit to participant ratings ($r(90) = 0.759$). This confirms that the model provides a decent first pass at predicting how speakers evaluate statistical trends concerning gradient phonotactic probability given the data of the language.

3.4. Testing for noun vs. verb differences

The result in Figure 3 assumes that statistical probabilities are calculated over the lexicon as a whole without regard for part of speech. We are now in a position to address the question of interest, which is whether a more accurate model of speakers' intuitions can be obtained by modeling nouns and verbs separately. In order to do this, the model was trained twice, first on the noun lemmas in CELEX, and then on the verb lemmas. These models were then used to derive predictions for the 375 novel words used in the experiments described above. As noted above, not all novel items contained features that would encourage different ratings depending on their part of speech—a fact which is also reflected in a very high correlation between the noun and verb models (experiment 1: $r(168) = .972$; experiment 2: $r(203) = .974$). Therefore, our first task is to isolate the set of words that are most likely to yield an effect. The predictions of the noun and verb models were correlated against one another, and for each test item the difference between the predictions of the two models was calculated (i.e., the residuals). We can take the absolute value of this difference as a metric of how large of an effect we might expect for each particular novel word. We can take the sign of the difference as a predictor of whether the word should sound better

[5] I ignore here a normalization term that is needed to convert this product into a true probability.

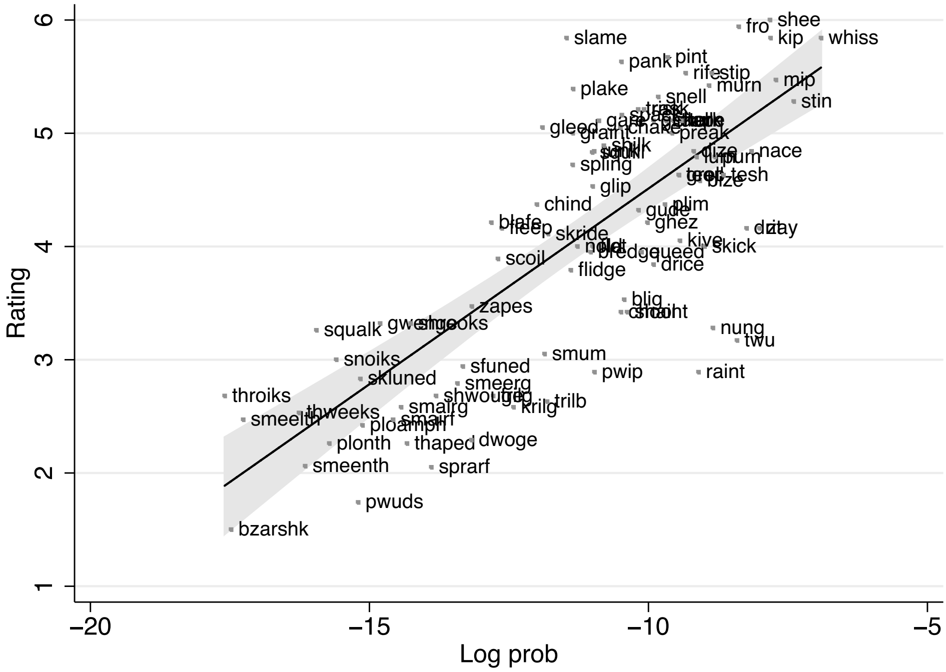

Figure 3: Performance of the model on 92 novel words (all presented as verbs)

as a noun or as a verb. For example, the novel words *ict* [ɪkt] and *reeze* [riːz] are predicted to sound better as verbs than as nouns, since they contain clusters or final segments that are commonly found in verbs. The words *grelm* [grɛlm] and *chake* [t͡ʃeɪk], on the other hand, have higher scores in the noun model since they contain onsets and codas that are not found especially often among verbs.

By focusing on the words with the greatest noun/verb discrepancy, it is possible to isolate sets of items that reduce the predicted correlation between noun and verb ratings. As it turns out, for experiment 1 the most divergent set consisted of 21 items, with a predicted noun/verb correlation of $r = .88$ (down from .97); for experiment 2 the most divergent set included 56 items, with a predicted correlation of .90 (also down from .97). Although these sets of words by no means eliminate the correlation, they maximize our chances of finding noun-verb differences in the acceptability ratings.

In order to test whether English speakers learn separate phonotactics for nouns vs. verbs, ratings of noun and verb presentations were compared against all three models: the “baseline” model trained on lemmas representing all parts of the speech, the noun model trained specifically on noun lemmas, and the verb model trained just on verb lemmas. If speakers learn separate probabilities for phonotactic rules/rankings depending on the class of the word, then the best model of acceptability ratings should be one that is trained specifically on the same class that was used in the experimental presentation (that is, the noun model for nouns, the verb model for verbs). In (7), we see that in fact models trained specifically on lemmas of the same class do outperform the baseline (all lemmas) model, although in some cases the improvement is quite modest.

(7) Modeling gain from using class-specific models:

		Baseline model	Same-class model	Improvement
Verb presentations	Exp 1 ($n = 21$)	.474	.502	+.028
	Exp 2 ($n = 56$)	.460	.486	+.026
Noun presentations	Exp 1 ($n = 21$)	.456	.478	+.022
	Exp 2 ($n = 56$)	.427	.430	+.003

On the face of it, the improvements in the last column of (7) provide weak but positive evidence for the idea that speakers learn distinct statistical tendencies for nouns vs. verbs. This is consistent with the hypothesis that bottom-up learning always starts with narrow, class-specific comparisons. Support for this is diminished, however, if we consider the performance of just the verbs-only model. As the table in (8) shows, using a verbs-only model improves the results not only for verb presentations (repeated from (7)), but surprisingly also for noun presentations. In other words, the best model of acceptability ratings is not one in which novel words are compared to existing words of the same part of speech, but rather, it is one in which novel words are compared only to existing *verbs*, regardless of the part of the speech.

(8) Modeling gain from using a verbs-only model for both nouns and verbs:

		Baseline model	Verbs-only model	Improvement
Verb presentations	Exp 1 ($n = 21$)	.474	.502	+.028
	Exp 2 ($n = 56$)	.460	.486	+.026
Noun presentations	Exp 1 ($n = 21$)	.456	.535	+.079
	Exp 2 ($n = 56$)	.427	.451	+.024

This result is unexpected, and may seem a bit illogical: why would ratings of nouns be better modeled by a training set of verb lemmas? It seems likely that this does not actually reflect a greater primacy of verbs in defining phonotactic grammar. Rather, it appears that this result may be a side-effect of the fact that the CELEX verb lemmas represent a smaller training set (5257 items, compared with 21571 nouns), with fewer odd or rare sequences. The relative lack of marginal sequences among verbs may be due in part to sampling: there are simply fewer verbs, so fewer opportunities for rare sequences to show up among them. Another important contributing factor may be the fact that borrowings into English tend to be nouns, so there is greater potential to introduce unusual sequences into the set of noun lemmas (*concierge* [ɛɪʒ], *bwana* [#bw], etc.).[6] It appears that CELEX counts may not adequately differentiate very rare/borrowed sequences from somewhat rare ones—perhaps reflecting the fact that the corpus is based at least in part on newspapers and written text, in which distant and exotic themes may be statistically overrepresented. A similar effect has been observed by Hayes and Wilson (in press), who find that their models achieve better performance on English onset clusters when they use a training corpus that is purged of onsets deemed to be foreign or unusual. Hayes and Wilson speculate that such words may have combinations of properties that mark them as “non-English” (many unusual properties clustered in a single word, or referring to exotic objects), leading learners to ignore them for purposes of learning phonotactic grammars. Regardless of whether it is merely a sampling effect or some more interesting selectional effect on the part of learners, it appears that corpora may sometimes overestimate the goodness of unusual sequences, and models that exclude them in one way or another tend to perform better. We may surmise that the verbs-only model does better overall for this reason, and not because English speakers restrict themselves to verbs when learning or assessing phonotactic well-formedness.

The upshot is that we do not observe an effect of class-specific phonotactics in English.[7] As a final test of whether any consistent noun/verb differences can be found in isolated corners of the data, we may

[6]The relatively greater importance of loans as opposed to sampling error may be seen from the fact that some rare native sequences actually occur more often in verbs than in nouns: [θw] *thwart* (v.), *thwack* (v./n.); [dw] *dwell* (v.), *dwindle* (v.), *dwarf* (n./v.).

[7]Analyses carried out over the entire set of items, rather than specific subsets, yielded essentially the same results but in even stronger terms: a model based only on verbs consistently outperforms the baseline model for both parts of speech, while a model based only on nouns sometimes does even worse than the baseline in modeling nouns.

also compare ratings for items that contained just those features that have been discussed in the literature: final fricative voicing and backness of the stressed vowel. (Recall that we cannot test for stress or syllable count differences, because only monosyllabic nonce words were tested.) For each test item, the degree of "noun/verb preference" was calculated as follows: first, the experimentally obtained ratings of verb presentations were fit to the ratings of noun presentations by linear regression. As above for the outputs of the models, the residuals were then calculated, in order to provide an estimate of whether the verb rating was higher or lower than expected based on the corresponding noun rating. If the result is positive, this indicates that the word was preferred as a verb; if the result is negative, the word was preferred as a noun.

What we would like to know is whether words ending in voiced fricatives tend to have greater or lesser noun preference than words ending in voiceless fricatives (and correspondingly, for front vs. back vowels). Therefore, the noun/verb preference scores were submitted to one way ANOVA's, testing for effects of fricative voicing and vowel backness.[8] The results showed no significant effects, either for vowel backness (experiment 1: $F(1) = 2.22$, $p = .139$; experiment 2: $F(1) = 0.73$, $p = .395$) or for fricative voicing (experiment 2: $F(1) = 0.00$, $p = .969$). This result helps to confirm the conclusion above that the lack of improvement from using class-specific training sets is not merely due to the fact that the model is not an adequate model of gradient differences. Rather, it appears that the expected differences between nouns and verbs are simply not found in participants' ratings.

Naturally, this negative result leads us to wonder whether the manipulation failed, and participants did not pay attention to the part of speech when listening to the novel words. There is anecdotal evidence suggesting that this is not the case. First, a number of subjects reported during debriefing that they had not really been able to judge words based on how they sounded, and that instead they had based their ratings of individual items mainly on their part of speech. The high degree of correlation between noun and verb ratings and the lack of interpretable differences between the two shows that this assessment cannot be right; in fact ratings were based almost exclusively on the segmental composition of the words and depended rather little on their part of speech. However, the fact that many subjects reported paying attention to the noun/verb difference shows that it was not being ignored completely.

Further evidence showing that subjects did pay attention to and process the part of speech comes from another phenomenon that emerged sporadically in the responses: aggressive overparsing.[9] Some of the novel items ended in clusters that most frequently arise through affixation: *neepse* [niːps], *glact* [glækt], and so on. In such cases, subjects sometimes erroneously removed the final consonant when repeating the word: "[niːps]. I like to [niːp]". This confirms that subjects were not listening to novel items purely as strings of segments and disregarding syntactic information. Instead, the occurrence of overparsing shows that test items were represented with morphological and syntactic structure, which in principle could have played a role in their acceptability ratings.

The provisional conclusion, then, is that participants in these experiments did pay attention at some level to the noun/verb status of novel items, but that this information does not enter into the calculation of phonotactic well-formedness. This supports a model in which segmental phonotactics are not relativized to word class in English—or, at least, not to the noun/verb distinction.[10] If this is correct, it puts us in a bit of a quandary, since it goes against a considerable literature documenting ways in which noun/verb differences are actively enforced by speakers (e.g., by fricative voicing) and play a role in a variety of behaviors in experimental tasks. I will argue here, however, that none of the observed effects actually require us to assume that the phonotactic grammar of English is divided by part of speech.

The first type of data that must be explained are the voicing alternations in pairs such as *hou*[s]*e* ∼ *hou*[z]*e* and *ba*[θ] ∼ *ba*[ð]*e*. In theory, one might be tempted to attribute such alternations to class-specific phonotactic constraints, such as "no voiceless fricatives at the ends of verbs" ((9a)). This constraint, ranked probabilistically in such a way that voiceless fricatives are at least sometimes banned at the ends of verbs (Boersma & Hayes 2001; Zuraw 2000), would be sufficient to produce a semi-productive effect

[8]The two features were tested in separate ANOVA's, since for experiment 1 words, there were not enough items ending in voiced fricatives to test this factor.

[9]I borrow the term AGGRESSIVE from Hammond (1999) and Zuraw (2002), who use it to refer to cases in which morphological structure is imposed although it is not supported by meaning or syntax.

[10]Systematic differences between function and content words (e.g., *[#ð]) seem plausible, but are harder to test because it is difficult to make up novel function words. For one attempt, with suggestive results, see Campbell & Besner (1981).

of fricative voicing in verbs, as in (9b).

(9) a. No verb-final voiced fricatives: $* \begin{bmatrix} -\text{son} \\ +\text{cont} \\ -\text{voi} \end{bmatrix} /\_\_\_\,]_{\text{Verb}}$

b. Preference for [z] in *house* (verb)

/haʊs/$_{\text{Verb}}$	$* \begin{bmatrix} -\text{son} \\ +\text{cont} \\ -\text{voi} \end{bmatrix} /\_\_\_\,]_{\text{Verb}}$	MAX[−voi]
a. haʊs	*!	
☞ b. haʊz		*

The fact that we cannot observe any independent effect of constraint (9a) casts doubt on such an analysis. While there does not appear to be any general phonotactic dispreference for verbs to end in voiceless fricatives, there is certainly a preference for derivationally related nouns and verbs to differ in voicing, with noun-final voiceless fricatives corresponding to verb-final voiced fricatives. It is possible to encode this relation directly in the morphological conversion process, as in (10).

(10) Conversion with voicing:

$$\text{X} \begin{bmatrix} -\text{son} \\ +\text{cont} \\ -\text{voi} \end{bmatrix}]_{\text{Noun}} \rightarrow \text{X}\ [+\text{voi}]\ ]_{\text{Verb}}$$

Under this analysis, voicing is not motivated by phonotactic pressure, but may rather be viewed as part of the morphological exponence of denominal verbs.[11] Stress alternations in pairs like *récord* vs. *recórd* could be handled in similar fashion if it turned out that speakers are equally likely to accept novel words as nouns or verbs regardless of the stress pattern. By shifting the analysis of alternations in related pairs out of the phonology and into the morphology in this way, we are able to reconcile the preference for voicing alternations specifically in related pairs, but not in underived forms.[12]

The current result also appears to stand against a sizeable literature showing processing differences between nouns and verbs (see, e.g., Kelly 1992). Most studies have focused on stress differences, showing that stress can influence the assignment of unknown words to categories ("is *blick* a noun or a verb?"), willingness to apply inflection or morphological conversion ("which sounds better as a verb, 'to *otter*' or 'to *raccoon*'?"), and so on. While segmental differences have not been tested as extensively as stress, they have also occasionally been implicated in reaction time differences in lexical decision tasks (Sereno & Jongman 1990; Sereno 1994; Farmer *et al.* 2006). However, these studies all involve tasks that are quite different from the current experiment, which simply asks how acceptable a given string is as a member of category X. We know that tasks like lexical decision or guessing the category of an unknown word tap many kinds of knowledge beyond phonotactic well-formedness, including token frequency, neighborhood density, context, and others. It is well known that word recognition is strongly influenced by the existence of lexical neighbors (Luce 1986), so it is not surprising if a high proportion of same-category neighbors is particularly beneficial in recognizing words in syntactic context (Farmer *et al.* 2006). Likewise, decisions about whether an unknown word is more likely to be a noun or a verb may depend not only on its phonotactic probability, but also on the level of activation of similar noun/verb neighbors. In short, it appears that previous studies demonstrating interactions between phonological form and part of speech have employed tasks that involve polling the lexicon (what Schütze 2005 calls "dictionary scenarios"). If we assume a model in which knowledge of sequence probability (phonotactics) is distinct from knowledge of existing words (the lexicon), it is completely possible that phonotactic knowledge is not relativized to nouns vs. verbs while lexical access is (perhaps necessarily) sensitive to this distinction. In short, previous work showing behavioral differences between nouns and verbs based on their segmental make-up do not require separate phonotactic grammars for nouns and verbs.

[11] A similar view is advocated by Chomsky & Halle (1968:232), who claim that voicing adjustments are "associated with" derivational operations.

[12] It would be desirable to test whether the strength of the voicing effect is accurately predicted by the number of attested noun/verb pairs that differ in voicing, perhaps by testing the productivity of the alternation on nonce words.

To sum up the results of this section, it appears that phonotactic acceptability judgments in English do not depend on the part of speech, in spite of the fact that there are observable statistical differences between verbs and nouns. Although this conclusion rests in part on a null result, it is accompanied by the positive claim that the best model of nonword ratings, regardless of their part of speech, is a relatively small, "clean" set of existing words (embodied here by the set of verb lemmas in CELEX). I have argued further that subjects did not ignore the part of speech completely in the experiments reported here, but rather, they simply did not use it to inform their ratings. This supports the claim that the grammar of English phonotactics is not relativized to specific parts of speech. This conclusion has implications for the analysis of a number of phenomena in English which distinguish between nouns and verbs. The more important issue for present purposes, however, is to understand why English speakers are insensitive to gradient statistical differences between word classes, while Spanish speakers maintain such differences to the point where they are reluctant to generalize across classes even when there is no other data available. In the remainder of this paper, I discuss an approach that seems helpful in resolving this contradiction.

4. Discussion

Comparing the data from the previous two sections, we arrive at a paradox: on the one hand, the best model of Spanish vowel phonology is one that treats all patterns as conjugation class specific, even when such restrictions are unmotivated and impede inflection. In particular, diphthongization of /o/ under stress is not generalized from classes 1 and 2 to class 3, even though class 3 has practically no existing /o/ verbs and therefore cannot have an observably different rate of diphthongization than the other two classes. The result of this restriction is that speakers are unable to decide whether or not to diphthongize unknown class 3 words, and are stuck with a paradigm gap (Albright 2003). The restriction of /o/ diphthongization to individual conjugation classes is a challenge to the idea that learners posit class-by-class differences only in the face of positive evidence. On the other hand, the best model of English segmental phonotactics is one that ignores word class, in spite of the fact that there are observable statistical differences between nouns and verbs. In this case, positing phonotactic differences could actually help speakers make sense of alternations such as *hou*[s]*e* (n.) ∼ *hou*[z]*e* (v.), which instead must be treated as arbitrary morphological adjustments. The lack of sensitivity to noun/verb differences is puzzling under a purely bottom-up account in which detailed class-specific knowledge is an automatic consequence of inductive learning. We therefore see "poor fits" in both directions: cases where phonology is indexed to particular classes in the absence of positive evidence, and cases where it is not indexed in spite of statistical differences.

The claim of this paper is that such mismatches can be understood by looking at the language more broadly. Although there is no evidence for class-by-class differences in /o/ diphthongization in Spanish, there is abundant evidence for class-by-class differences with respect to other properties: /e/ diphthongization, /e/ raising, velar insertion, the distribution of vowel height all differ substantially across the three conjugation classes. Conversely, although it is possible to observe noun/verb differences in features like final fricative voicing in English, these differences are small in magnitude and few in number compared to the large number of properties that the classes have in common. This suggests that the decision about whether to subdivide phonological grammar in order to capture class-by-class differences may be a global one rather than a process-by-process choice. This is most compatible with a model in which indexation to particular classes is done at the level of the entire grammar (co-phonologies whose rankings must be learned separately), rather than at the level of the individual constraint.

As has been pointed out in the literature, the use of completely separate grammars for different sets of words can be quite uneconomical: it fails to capture systematic and seemingly significant parallels between the phonologies of different word classes (Itô & Mester 1995), and it has the danger of leading to redundant and needless proliferation of grammars for patterns that are shared across different classes (Inkelas *et al.* 1997). The decision to posit distinct grammars for different classes of words surely comes at a cost, which learners must weigh against the potential payoff in more accurately capturing the linguistic data. In this section, I sketch a (as yet unimplemented) model of how learners might assess this trade-off to select relativized grammars in some cases but not others, which also attempts to explain why completely distinct co-phonologies might be a preferable choice over independently indexed constraints.

4.1. A hierarchical model

A potentially useful class of models for the current problem are HIERARCHICAL MODELS (Good 1980; Kemp *et al.* 2007). The intuitive idea behind a hierarchical model is that learners seek to explain data by selecting hypotheses at multiple levels of abstraction. For example, faced with the task of learning phonology, learners might consider not only hypotheses about the correct constraint ranking (the usual definition of learning in OT; Tesar & Smolensky 2000), but also hypotheses about what the right constraints are, about what kinds of constraints should be considered, and perhaps even about whether a grammatical explanation of the data is appropriate in the first place. In the case of Spanish vowel phonology, learners must decide whether to posit a grammatical analysis or whether to simply list every form (lexical analysis). For the class 1 mid vowel verbs, a purely lexical analysis would require listing hundreds or thousands of verbs, most of which behave alike (default non-alternation). In this case, we have fairly strong evidence that speakers choose instead to learn a grammar of abstractions about the contexts that favor alternation or default non-alternation. For class 3, on the other hand, there are very few examples—particularly for the back vowel /o/ (*dormir*, *morir*, *oír*)—and it is not at all clear that speakers learn a grammar for this set of words.[13] Among the space of possible grammatical analyses, there are those that employ fully general constraints (*[+high], *[−high,−low]), there are those that employ relativized constraints ($*[+\text{high}]_{\text{Class 2}}$, $*[-\text{high},-\text{low}]_{\text{Class 3}}$), in addition to those that employ mixes of both types ($*[+\text{high}]_{\text{Class 2}}$, *[−high,−low], etc.). Furthermore, given a particular template for possible constraints, there are many different hypotheses about what the relevant constraints might be ($*[+\text{high}]_{\text{Class 2}}$, $*[+\text{high}]_{\text{Class 3}}$, etc.). Finally, there is a set of hypotheses about what ranking of constraints is needed to derive the observed data.

Clearly, the space of possible hypotheses can be enormous, since the learner has the freedom to choose not only the analysis (the particular ranking of constraints), but also the appropriate type of analysis. Such a model is less restrictive than the standard conception of OT, in which the hypothesis space is limited to hierarchies (rankings) of a pre-defined and universal constraint set. At the same time, the model has a considerable amount of structure built in: the rules for defining possible constraints, the syntax of possible kinds of constraints, and so on. Furthermore, learners may bring with them biases for certain types of analyses over others. These biases may include straightforward preferences for simpler analyses over more complicated analyses (e.g., a single grammar of general constraints rather than multiple grammars of class-specific constraints), and they may also include substantive biases for some constraints or constraint types over others (e.g., *k/___ [−back] rather than *k/___ [−high] ; Wilson 2006). Ultimately, the choice of a particular hypothesis rests on a combination of prior biases and the ability of the hypothesis to explain the data. (See Kemp *et al.* 2007 for an overview of hypothesis selection in a Bayesian framework.)

The question of interest here is what considerations might prompt a learner to select class-specific grammars rather than a fully general one. In terms of accuracy, it would almost always be advantageous to select class-specific grammars, since they allow us to capture whatever statistical differences there are between words in different classes. What we need is some way of encoding the cost associated with selecting a relativized grammar. In a Bayesian model, this comes in the form of a lower prior probability.

Let us assume that in a hierarchical model of grammar construction, there is a level at which learners hypothesize what kinds of constraints are involved. The space of possible hypotheses might include fully general constraints, or constraints relativized to a particular part of speech: function vs. content words, nouns vs. adjectives vs. verbs, and presumably a number of other possibilities, including intermediate subgroupings (verbs and adjectives vs. nouns, etc.). Constraints could also be relativized to different inflectional classes, or even to arbitrarily defined sets of words. Furthermore, grammars might consist of mixed constraint types: some constraints fully general and others relativized to particular classes. Some of these options introduce much larger numbers of hypotheses than others: there are more ways to construct grammars using combinations of general and relativized constraints than there are ways to define grammars with just one constraint type, and there are vastly many more arbitrarily defined sets of words than there are parts of speech.

Tenenbaum (2000) proposes that one source of bias for simpler or more restrictive types of

[13] See Tenenbaum (2000) for parallel non-linguistic examples contrasting exemplar- vs. rule-based generalization.

hypotheses may come from the fact that they come from "less crowded" parts of the hypothesis space. Suppose that the learner has no prior bias for one type of grammar over another, and is equally like to choose a single fully general grammar, a set of subgrammars relativized to specific word classes, or some mix of the two; that is, each type of grammar has a prior probability of 1/3. Since there many more ways to construct relativized grammars, there are more hypotheses of this type competing for a share of the 1/3 of the probability mass allotted to relativized grammars. The suggestion, then, is that at least some prior biases may come from the way that the hypothesis space is structured, with probability mass distributed equally at higher levels in the hierarchy translating into unequal prior probabilities of individual low-level hypotheses.

In the current case, this appears to make the correct prediction. The prior probability of unified grammars would be highest, and the learner would be biased against relativized grammars; grammars that mix relativized and unrelativized constraints would have the very lowest prior probability. Crucially, in a Bayesian model it is often possible to favor hypotheses with low prior probability, provided that they fit to the data very accurately. For example, a grammar of Spanish that encodes class-by-class differences in the distribution of vowel heights and alternations is much more accurate than a grammar that ignores conjugation class, because of the substantial statistical differences between the classes. When the accuracy payoff is small as in English, however, the difference in prior probability may be sufficient to favor a simpler but slightly less accurate analysis that ignores conjugation class.

This line of analysis also makes an intriguing prediction concerning "mixed" grammars containing both general and indexed constraints. It is always possible to recast a mixed grammar as a purely relativized grammar, since the general constraints can always be split (inefficiently) into their class-specific variants. Therefore, it is never possible for a mixed grammar to be more accurate than its corresponding completely relativized grammar. As mentioned above, the need to redundantly restate general properties separately for each word class is often taken to be a liability for analyses with completely separate subgrammars. However, when we consider not just the number of constraints involved but also the complexity of specifying those constraints, a purely relativized grammar may actually be simpler, since we know that for every constraint we must specify (1) the word class and (2) the phonological condition. In order to specify a constraint in a mixed grammar, on the other hand, we must specify (1) whether it is general or specific, (2) if specific, what class it refers to, and (3) the phonological condition.[14] Under the plausible assumption that larger hypothesis spaces imply smaller prior probability for any individual low-level hypothesis, mixed grammars would be at an insurmountable disadvantage, since they can never achieve the greater accuracy that is needed to overcome their lower prior probability. In other words, the preference for "all or nothing" relativization may not be due to an absolute restriction on what speakers can encode, but may instead emerge as a side-effect of how hypotheses are evaluated.

This same mechanism may also favor grammatical distinctions between certain types of word classes over others. Assuming that there is a small universal set of lexical categories (noun, verb, etc.), the number of ways to relativize constraints according to syntactic categories is quite small. The number of inflectional classes in a language is often larger, and of course the number of possible arbitrarily defined sets of words is vast. This may result in a natural bias towards constraints indexed to certain types of word classes, such as nouns vs. verbs, by virtue of the fact that they compete with fewer other hypotheses of the same form.

Naturally, the analysis sketched here is quite speculative in the absence of a computationally implemented model that actually defines the hypothesis space, assigns prior probabilities over different hypothesis types, and selects hypotheses in response to language data. My goal in this section has been to suggest a mode of analysis that seems helpful in resolving the apparent paradox of why learners appear to settle on analyses that are sometimes more specific and at other times more general than the data demands. If this approach is on the right track, it has the potential to capture the way in which clear differences between word classes for some properties and some classes (e.g., the rate of high vowels in Spanish class 1 vs. class 2) can force a class-specific analysis of other properties, even in the absence of clear positive evidence. Conversely, it can explain why learners of a language with relatively few class-by-class differences would favor hypotheses in which there are never such differences, even if a few such differences can in fact be observed.

[14] Put differently, mixed grammars are just relativized grammars with one more possible specification for constraint indices: "all classes".

4.2. Alternations vs. static phonotactics?

Strictly speaking, the Spanish and English cases discussed here are not a perfectly matched comparison. In Spanish, the strongest evidence for unmotivated lack of generalization comes from morphophonological alterations (diphthongization of /o/). In English, by contrast, the data concerns static phonotactic distributions, which are perhaps not sensitive to all of the same morphological and syntactic conditioning. It is commonly assumed in the OT learning literature that children may learn a good deal about static phonotactic distributions prior to the point when they begin to acquire word meanings, parse words into morphemes, and compare related forms to discover alternations (Hayes 2004; Prince & Tesar 2004; Hayes and Wilson, to appear). If this is true, then it would mean that the noun/verb distinction may be irrelevant (or only partially available) at the time when contrasts in vowel backness and fricative voicing are being learned. This may indeed provide a partial alternative account of the Spanish/English difference. However, there are reasons to think that it cannot be the whole story.

The first relevant fact is that conjugation class differences in Spanish influence not only alternations such as diphthongization and vowel raising, but also static phonotactic distributions of features such as vowel height (Figure 1) and root-final voicing (Davis & Napoli 1994). If learning of phonotactic distributions was completed prior to any significant amount of morphological learning, then vowel height and stem-final voicing in Spanish should behave like fricative voicing in English, and not be sensitive to morphological class. The fact that at least some phonotactic distributions are sensitive to conjugation class in Spanish indicates that the difference cannot be pinned completely on alternations vs. static distributions.

There is also a suggestive converse fact concerning alternations in English. In standard descriptions of English, fricative voicing is said to alternate in two contexts: in the inflected forms of certain nouns (*knife* [naɪf] ∼ *knives* [naɪvz]; *house* [haʊs] ∼ *houses* [haʊzəz]; *path* [pæθ] ∼ *paths* [pæðz]), and in certain noun/verb pairs, as discussed above (*house* [haʊs] ∼ *house* [haʊz]; *bath* [bæθ] ∼ *bathe* [beɪð]; etc.). In this data, which reflects the historical context for fricative voicing, nouns and verbs show a difference: noun paradigms have alternation within the paradigm (unsuffixed = voiceless, suffixed = voiced) while verb paradigms are voiced throughout. However, it appears that for at least some speakers (myself included), it is marginally possible to extend the voicing alternation to verb paradigms:

(11) Unsuffixed: preference for voiceless

- a. ✓ I guarantee this sentence will garden path [pæθ] you.
- b. *? I guarantee this sentence will garden path [pæð] you.
- c. ✓ He'll sneak up and knife [naɪf] you in the back.
- d. *? He'll sneak up and knive [naɪv] you in the back.

(12) Suffixed: optionally (or preferably?) voiced

- a. ✓ That sentence has garden pathed [pæθt] me many times.
- b. ✓ That sentence has garden pathed [pæðd] me many times.
- c. ✓ He snuck up and knifed [naɪft] me in the back.
- d. ✓ He snuck up and knived [naɪvd] me in the back.

The difference is easiest to observe for final [f]/[v], since this distinction is represented orthographically and can be investigated in corpus searches. The denominal verb 'to knife' standardly maintains voiceless [f], in violation of the semi-productive tendency to voice derived verbs. Google searches show that as with most verbs, the present tense form of this verb is more frequent than the past: for example, in searches conducted on June 29, 2007 from Cambridge, MA, the phrase '*knife him*' elicited approximately 15,000 hits, while '*knifed him*' returned approximately 10,700 hits. There are approximately 50 hits for '*knive him*' with (orthographic) voicing, compared with 77 hits for '*knived him*', indicating a greater degree of voicing in the past. Similarly there are more hits for the past tense '*vouchsaved*' than for a present with voicing '*vouchsave*' (ca. 200 vs. 50). Although the numerical differences are somewhat weak, they point in the same direction as the intuitions in (12), in that the relative preference for voicing is reversed between the two morphological contexts (unsuffixed vs. suffixed). Crucially, this apparent willingness to take a pattern of alternation previously restricted to nouns and extend it to verbs indicates that alternations are not intrinsically limited to specific word classes, any more than static distributions are.

4.3. Parts of speech vs. inflectional classes

Another obvious way in which English differs from Spanish is that in the former case we are dealing with phonological differences between different parts of speech, while in the latter case, we are dealing with differences between different conjugation classes. Impressionistically, it appears that phonological differences between semantically arbitrary inflectional classes are quite common (characteristic, even), while systematic segmental differences between, say, nouns and verbs are relatively less robust. Although I know of no concrete attempts to make this comparison across a variety of languages, it is not terribly surprising, if true. Recall from the end of section 3.4 that one experimental condition in which English speakers do distinguish nouns from verbs is when they are forced to decide the more likely part of speech of an unknown word. I argued that this task does not necessary require a grammatical evaluation of well-formedness, but rather a comparison to the set of existing lexical items in order to classify the novel item. If learners were frequently faced with the task of deciding the category of unknown words based only on their phonological form, they would tend to arrive at lexicons in which there is a strong correlation between phonological form and word class. In the real world, there are contextual cues (semantic and syntactic) to the part of speech of a word, and if the learner makes a mistake when first encountering the word, they are likely to encounter it in a less ambiguous situation the next time. Inflectional classes, on the other hand, provide very little in the way of external cues to their membership, since they are not distinguished by their syntactic context and (unlike gender) they generally do not trigger morphological agreement. Furthermore, inflectional classes are often neutralized with one another, making ambiguity rampant in the input data. Therefore, it seems likely that learners have far more opportunity to impose their preferred phonology/category correspondences on inflectional classes than on part of speech. If this is right, then it provides an external account of why inflectional classes and parts of speech might show such different rates of phonological differentiation, and we do not need to build in any further mechanism to explain the difference.

5. Conclusion

To summarize, in this paper we have seen a two-sided puzzle. In Spanish, speakers fail to generalize vowel contrasts and alternations across different conjugation classes. This is true even for cases in which there is no overt evidence, such as the behavior of /o/ under stress in the third conjugation. In English, on the other hand, speakers systematically fail to distinguish among classes, even for properties that do show statistically significant differences. I have claimed that these seemingly suboptimal decisions about the scope of phonological constraints may follow from a more general preference for consistent constraint types—either all morphologically restricted, or none. In both Spanish and English, the choice of a restricted vs. unrestricted grammar makes sense given the system as a whole. I have further argued that this "global" nature of the restriction is appropriately captured by a hierarchical model in which choosing a hypothesis about the right type of constraints is an explicit part of the learning task. The next step is to provide a computational implementation of this idea as a hierarchical Bayesian model, in order to make good on the idea that "all or nothing" behavior could emerge as a result of the trade-off between simple and accurate theories. Finally, a great deal more empirical work is needed for these and other languages, in order to test the predictions of such a model for linguistic data with different kinds and degrees of phonological differences between word classes.

References

Albright, Adam. 2002. Islands of reliability for regular morphology: Evidence from Italian. *Language* 78, 684–709.

——. 2003. A quantitative study of Spanish paradigm gaps. In *WCCFL 22 Proceedings*, ed. by G. Garding & M. Tsujimura, 1–14, Somerville, MA. Cascadilla Press.

——, in prep. Gradient phonological acceptability as a grammatical effect. MIT ms.

——, Argelia Edith Andrade, & Bruce Hayes. 2001. Segmental environments of Spanish diphthongization. In *UCLA Working Papers in Linguistics, Number 7: Papers in Phonology 5*, ed. by Adam Albright & Taehong Cho, 117–151. http://www.linguistics.ucla.edu/people/hayes/Segenvspandiph/SegEnvSpanDiph.pdf.

——, & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90, 119–161.

Anttila, Arto. 2002. Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* 20, 1–42. Available for download at http://roa.rutgers.edu/view.php3?roa=425.

Baayen, R. H., R. Piepenbrock, & H. van Rijn. 1993. *The CELEX lexical data base on CD-ROM*. Philadelphia, PA: Linguistic Data Consortium.

Bailey, Todd, & Ulrike Hahn. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44, 568–591.

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

Berg, Thomas. 2000. The position of adjectives on the noun–verb continuum. *English Language and Linguistics* 4, 269–293.

Berko, Jean. 1958. The child's acquisition of English morphology. *Word* 14, 150–177.

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

Boyé, Gilles, & Patricia Cabredo Hofherr, 2006. The distribution of prethematic vowels in Spanish verbs. ERSS ms.

——, & Fabio Montermini, 2007. Regularity and irregularity in Italian verbal inflection. Poster presented at the 15th Manchester Phonology Meeting, May 2007.

Brame, Michael K., & Ivonne Bordelois. 1973. Vocalic alternations in Spanish. *Linguistic Inquiry* 4, 111–168.

Burzio, Luigi. 1994. *Principles of English stress*. Cambridge: Cambridge University Press.

Butt, John. 1997. *Spanish Verbs*. Oxford University Press.

Bybee, Joan, & Carol Moder. 1983. Morphological classes as natural categories. *Language* 59, 251–270.

Bybee, Joan L., & Elly Pardo. 1981. On Lexical and morphological conditioning of alternations: a nonce-probe experiment with Spanish verbs. *Linguistics* 19, 937–968.

Campbell, R., & D. Besner. 1981. This and thap: Constraints on the pronunciation of new written words. *Quarterly Journal of Experimental Psychology* 33, 375–396.

Carreira, María. 1991. The alternating diphthongs of Spanish: A paradox resolved. In *Current Studies in Spanish Linguistics*, ed. by Héctor Campos & Fernando Martínez-Gil. Washington, D.C.: Georgetown University Press.

Cassidy, Kimberly Wright, & Michael H. Kelly. 1991. Phonological information for grammatical category assignments. *Journal of Memory and Language* 30, 348–369.

Chomsky, Noam, & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Coleman, J. S., & Janet Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. In *Computational Phonology. Third Meeting of the ACL Special Interest Group in Computational Phonology*, 49–56. Somerset, NJ: Association for Computational Linguistics.

Davis, Stuart, & Donna Jo Napoli. 1994. *A prosodic template in historical change: The passage of the Latin second conjugation into Romance*. Torino: Rosenberg and Sellier.

Eddington, David. 1998. Spanish Diphthongization as a Non-derivational Phenomenon. *Rivista di Linguistica* 10, 335–354.

——. 2000. Spanish stress assignment within Analogical Modeling of Language. *Language* 76, 92–109.

Farmer, T.A., M.H. Christiansen, & P. Monaghan. 2006. Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences* 103, 12203–12208.

Frisch, S. A., N. R. Large, & D. 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.

Garcia-Bellido, Paloma. 1986. Lexical diphthongization and high-mid alternations in Spanish: An autosegmental account. *Linguistic Analysis* 16, 61–92.

Good, I.J. 1980. Some history of the hierarchical Bayesian methodology. In *Bayesian statistics*, ed. by J.M. Bernardo, M.H. DeGroot, D.V. Lindley, & A.F.M. Smith, 489–519. Valencia: Valencia University Press.

Grimshaw, Jane, & Alan Prince, 1986. A prosodic account of the to-dative alternation. Brandeis University ms.

Guion, Susan G., J.J. Clark, Tetsuo Harada, & Ratree P. Wayland. 2003. Factors affecting stress placement for

English non-words include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech* 46, 403–427.

Hammond, Michael. 1984. *Constraining Metrical Theory: A Modular Theory of Rhythm and Destressing*. Indiana University Linguistics Club.

——. 1999. *The Phonology of English: A Prosodic Optimality-Theoretic Approach*. Oxford University Press.

Harris, James. 1969. *Spanish Phonology*. Cambridge, MA: MIT Press.

——. 1977. Remarks on diphthongization in Spanish. *Lingua* 41, 261–305.

——. 1985. Spanish diphthongisation and stress: A paradox resolved. *Phonology Yearbook* 2, 31–45.

——. 1992. *Spanish stress: The extrametricality issue*. Bloomington: Indiana University Linguistics Club.

——. 1995. Projection and edge marking in the computation of stress in Spanish. In *The Handbook of Phonological Theory*, ed. by John A. Goldsmith, 867–887. Oxford: Blackwell.

Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

——. 2004. Phonological acquisition in Optimality Theory: The early stages. In *Fixing Priorities: Constraints in Phonological Acquisition*, ed. by R. Kager, J. Pater, & W. Zonneveld. Cambridge, UK: Cambridge University Press.

——, & Colin Wilson. to appear. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* .

Hooper, Joan. 1976. *An Introduction to Natural Generative Phonology*. New York: Academic Press.

Inkelas, Sharon, Orhan Orgun, & Cheryl Zoll, 1997. Exceptions and static phonological patterns: Cophonologies vs. prespecification. ROA 124.

Itô, Junko, & Armin Mester. 1995. The core-periphery structure in the lexicon and constraints on reranking. In *Papers in Optimality Theory*, ed. by Dickey Beckman & Urbanczyk, volume 18 of *University of Massachusetts Occasional Papers*. Amherst, Mass: GLSA.

——, ——, & Jaye Padgett. 1996. Licensing and redundancy: Underspecification in Optimality Theory. *Linguistic Inquiry* 26, 571–614.

Jackendoff, Ray. 1997. Twistin' the night away. *Language* 73, 534–559.

Jespersen, Otto. 1942. *A Modern English Grammar on Historical Principles*, volume 6 (Morphology). London: George Allen & Unwin Ltd.

Kelly, Michael H. 1992. Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review* 99, 349–364.

Kemp, Charles, Amy Perfors, & Joshua B. Tenenbaum. 2007. Learning overhypotheses with hierarchical Bayesian models. *Developmental Science* 10, 307–321.

Kenstowicz, Michael, & Charles Kisseberth. 1977. *Topics in Phonological Theory*. New York: Academic Press.

Köpcke, K.-M. 1993. *Schemata bei der Pluralbildung im Deutschen. Versuch einer kognitiven Morphologie*. Number 47 in Studien zur deutschen Grammatik. Tübingen: Narr.

Liberman, Mark, & Alan S. Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8, 249–336.

Luce, P. A. 1986. Neighborhoods of words in the mental lexicon. Technical report, Speech Research Laboratory, Department of Psychology, Indiana University.

McCarthy, John. 2005. Optimal paradigms. In *Paradigms in Phonological Theory*, ed. by Laura J. Downing, Tracy Alan Hall, & Renate Raffelsiefen, 170–210. Oxford: Oxford University Press. ROA-485.

Morris, Richard E. 2005. Attraction to the unmarked in Old Spanish leveling. In *Selected Proceedings of the 7th Hispanic Linguistics Symposium*, ed. by David Eddington, 180–191. Somerville, MA: Cascadilla Proceedings Project. http://www.lingref.com/cpp/hls/7/paper1097.pdf.

Newman, Paul. 2000. *The Hausa Language: An Encyclopedic Reference Grammar*. Yale University Press.

Oltra Massuet, Maria Isabel, & Karlos Arregi. 2005. Stress-by-structure in Spanish. *Linguistic Inquiry* 36, 43–84.

Pater, Joe. To appear. Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In *Phonological Argumentation: Essays on Evidence and Motivation*, ed. by Steve Parker. London: Equinox.

Penny, Ralph. 2002. *A History of the Spanish Language*. Cambridge University Press, 2nd edition.

Prince, Alan, & Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell Publishing.

——, & Bruce Tesar. 2004. Learning phonotactic distributions. In *Fixing Priorities: Constraints in Phonological Acquisition*, ed. by R. Kager, J. Pater, & W. Zonneveld. Cambridge, UK: Cambridge University Press.

Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace.

Scholes, Robert J. 1966. *Phonotactic Grammaticality*. Janua Linguarum. The Hague: Mouton.

Schuldberg, Howard Kelly. 1984. Diphthongization in Spanish Verbs. *Hispanic Linguistics* 1, 215–227.

Schütze, Carson. 2005. Thinking about what we are asking speakers to do. In *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, ed. by Stephan Kepser & Marga Reis, 457–484. Mouton de Gruyter.

Sebastián, Núria, Fernando Cuetos, M. Antònia Martí, & Manuel F. Carreiras. 2000. *LEXESP: Léxico informatizado* *del español. Edición en CD-ROM*. Barcelona: Edicions de la Universitat de Barcelona (Colleccions Vàries, 14).

Sereno, Joan A. 1994. Phonosyntactics. In *Sound Symbolism*, ed. by L. Hinton, J. Nichols, & J. Ohala, 263–275. Cambridge: Cambridge University Press.

——, & Alard Jongman. 1990. Phonological and form class relations in the lexicon. *Journal of Psycholinguistic Research* 19, 387–404.

Sezer, Engin. 1981. The k/∅ alternation in Turkish. In *Harvard Studies in Phonology*, ed. by N. Clements, 354–382. Bloomington: Indiana University Linguistics Club.

Smith, Jennifer. 1999. Noun faithfulness and accent in Fukuoka Japanese. In *WCCFL 18*, ed. by Sonya Bird, Andrew Carnie, Jason D. Haugen, & Peter Norquest, 519–531, Somerville, MA. Cascadilla Press.

——. 2001. Lexical category and phonological contrast. In *Papers in Experimental and Theoretical Linguistics 6: Workshop on the Lexicon in Phonetics and Phonology*, ed. by Robert Kirchner, Joe Pater, & Wolf Wikely, 61–72. Edmonton: University of Alberta.

Tenenbaum, J. B. 2000. Rules and similarity in concept learning. In *Advances in Neural Information Processing Systems 12*, ed. by S. Solla, T. Leen, & K. Muller, 59–65. Cambridge: MIT Press.

Tesar, Bruce, & Paul Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Vitevitch, Michael, & Paul Luce. 1998. When words compete: Levels of processing in perception of spoken words. *Psychological Science* 9, 325–329.

——, & ——. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40, 374–408.

——, & ——. 2005. Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory & Language* 52, 193–204.

——, ——, Jan Charles-Luce, & David Kemmerer. 1997. Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech* 40, 47–62.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science: A Multidisciplinary Journal* 30, 945–982.

Zuraw, Kie. 2000. *Patterned Exceptions in Phonology*. UCLA dissertation.

——. 2002. Aggressive reduplication. *Phonology* 19, 395–439.

Proceedings of the 26th West Coast Conference on Formal Linguistics

edited by Charles B. Chang and Hannah J. Haynie

Cascadilla Proceedings Project Somerville, MA 2008

Copyright information

Proceedings of the 26th West Coast Conference on Formal Linguistics
© 2008 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 1-57473-423-2 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Albright, Adam. 2008. How Many Grammars Am I Holding Up? Discovering Phonological Differences between Word Classes. In *Proceedings of the 26th West Coast Conference on Formal Linguistics,* ed. Charles B. Chang and Hannah J. Haynie, 1-20. Somerville, MA: Cascadilla Proceedings Project.

or:

Albright, Adam. 2008. How Many Grammars Am I Holding Up? Discovering Phonological Differences between Word Classes. In *Proceedings of the 26th West Coast Conference on Formal Linguistics,* ed. Charles B. Chang and Hannah J. Haynie, 1-20. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1651.