Sublexical Phonotactics and English Comparatives*

Maria Gouskova & Suzy Ahn

New York University, University of Ottawa

Abstract

English comparative *-er* is said to be phonologically selective, but the restrictions are too complex to state elegantly in a subcategorization frame, and too lexically specific to capture in an emergence of the unmarked account. We argue that selection is really lexical, and that phonological generalizations emerge from phonotactic learning over sublexicons associated with the affixation rule. This theory is tested in a rating experiment that presents nonce words in an adjectival ("very wug") or a verbal context ("to wug"), and then asks people to rate the suffixed forms presented as either comparatives ("much wugger") or deverbal nouns ("a wugger"). People rate adjectives higher than verbs in bare forms, but the ratings are reversed for suffixed forms. The ratings furthermore depend not just on syllable count and stress position but on the overall well-formedness of the wugs with respect to the sublexical phonotactics of adjectives that combine with the comparative *-er* suffix.

1. Introduction

1.1 The phonology of English adjective gradation

English uses the suffix *-er* to form comparatives, but only for some of its adjectives: *fit/fitter*, *happy/happier*, but not *productive/*productiver*. The factors that make *er*-suffixation possible have been the subject of intense study.[1] These investigations identify several phono-

*This paper was originally written and distributed as a manuscript in 2016. We received valuable feedback from Robert Daland, Lisa Davidson, David Embick, Gillian Gallagher, Bruce Hayes, Itamar Kastner, Elliott Moreton, Yining Nie, Jochen Trommer, Jen Smith, Barbara Stiebels, and audiences of NYU's Morphology Research Group and CUNY, Johns Hopkins, Roots IV, UCLA, UPenn, the University of Leipzig, and the 2016 LSA in DC. This research was supported by NSF BCS-1224652 to the first author.

[1] Just to name a few: Pesetsky (1985), McCarthy and Prince (1986), Marantz (1988), McCarthy and Prince (1990), Poser (1992), Leech and Culpeper (1997), Lindquist (2000), Kytö and Romaine (2000), Graziano-King and Cairns (2005), Embick (2007), Boyd (2007), Hilpert (2008), Embick and Marantz (2008), Mondorf (2009), Bobaljik (2012), Matushansky (2013). Some of these works will be discussed in more detail in the following pages.

logical generalizations about adjectives that combine with *-er*: the adjectives tend to be monosyllabic (e.g., *big/bigger*) or disyllabic (e.g., *sílly/síllier*), and for disyllables, the second syllable is phonologically restricted. This syllable tends to be unstressed, and it may end in an approximant but not an obstruent or a consonant cluster: in American English, we find *yéllower* but not *?rápider*, *?moróser*, ?*inténser*, or *?pléasanter*. Trisyllables generally do not combine with the comparative suffix: **mílitanter, *abúsiver.* Various semantic, syntactic and pragmatic factors have also been shown to affect the likelihood of *-er* affixation, but the question we would like to ask in this article is about the nature of the phonological restrictions: how are they encoded grammatically? Our answer is that they are not encoded in the rules of affixation themselves. Affixation rules specify lexical but not phonological context; the phonological information about adjectives that combine with *-er* is extracted through probabilistic sublexical phonotactic learning (Gouskova et al. 2015, Becker and Allen 2015, Becker and Gouskova 2016). The resulting sublexical phonotactic grammar can then be used to decide whether the suffix *-er* can attach to novel items.

We tested this theory in an experiment with nonce words (wugs, Berko 1958), where we presented English speakers bare wugs in a verbal or an adjectival context (see (1)). We then presented the same wug with two homophonous *-er* suffixes, appearing in a nominalized or a comparative context. If comparative affixation depends only on general phonological factors, then there should not be a difference between deverbal and deadjectival *-er* forms; if, on the other hand, affixation is conditioned by morpheme-specific factors, deadjectival *-er* comparatives should be rated worse than deverbal nouns:

(1) *Our nonce word experiment*

a. Wug presented as verb: "I like to *fudáddow*"
b. Wug presented as adjective: "I am very *fudáddow*"
c. Wug affixed with nominalizer *-er*: "I am an amateur *fudáddower*"
d. Wug affixed with comparative *-er*: "But you are even *fudáddower*"

The ratings of Verb/Verb-er forms were about the same (*to fudaddow = a fudaddower*), but people rated the same wug higher when presented as an adjective than as a verb (*very fudaddow > to fudaddow*), and lower when the wug was affixed with the comparative *-er* than with the nominalizing *-er* (*a fudaddower > even fudaddower*). This suggests that affixation is not just a matter of general phonological factors. The phonological factors did play a role, however: monosyllabic wugs were rated as more acceptable than disyllables, which in turn beat trisyllables, and segmental content mattered as well. The phonological factors affected the ratings of adjectives differently than verbs, consistent with a sublexical phonotactic account.

1.2 Phonological generalizations about English comparative formation

We start here by reviewing some facts about English comparatives that any theory should be able to account for. The research on comparatives has identified a number of semantic, syntactic and stylistic factors that affect the use of *Adj-er* vs. *more Adj* forms, but we will

focus mainly on the phonological factors. In most cases, the generalizations we formulate hold all other things being equal—by other things, we mean semantics, syntax, and register.

1.2.1 Size and stress

Both formal and descriptive accounts prominently feature the following generalization about phonological size (Pesetsky 1979:pp. 10–11 et seq.):

(2) There are monosyllables and disyllables that combine with comparative *-er*, but trisyllables generally do not.

The restriction on size can also be understood as a restriction on stress. This is an insight due to Prosodic Morphology—a theory that reduces syllable-counting generalizations in morphology to the shape and position of metrical feet in the language (McCarthy and Prince 1986 et seq.). The *-er* suffix tends to attach to moraic trochees: monosyllables or disyllables with initial stress. McCarthy and Prince (1986) mark *obtúser* as ungrammatical, but *stúpider* as grammatical (see also Leech and Culpeper 1997). The stress restriction is somewhat more cumbersome to state than syllable counting, but it has been argued to be more accurate, at least for American English. We show in (3) some examples of comparatives that obey the size and stress generalization along with ones that end in the same consonants but violate the size restriction.

(3) *Examples of the stress/size generalization*

	Obey:			*Violate:*		
a.	sane	saner	σ́	humane	humaner	σσ́
b.	loose	looser	σ́	obtuse	obtuser	σσ́
c.	able	abler	σ́σ	probable	probabler	σ́σσ

Despite the strength of this generalization, there are occasional exceptions such as "políter" and "obscéner", which occur in the Google Ngrams corpus with some frequency (see §1.2.3). There is also a systematic exception for trisyllables whose first syllable is the prefix *un-*, e.g., *ùnháppier*—such words appear to require contradictory syntactic and phonological analyses.

1.2.2 Restrictions on the content of the last syllable

There are also restrictions on the segmental content of the final syllable, and these restrictions are more stringent in disyllables than in monosyllables. Pesetsky (1979) notes that final [i] makes suffixation likelier in disyllables, for example. One can find examples of second-syllable sonorants ([oʊ] in *yellower*, [l̩] in *subtler*, [ɹ̩] in *bitterer*), but obstruents are rare, as are consonant clusters. Thus, compare (4a–h) with their counterparts in (4i–p)—none of the italicized disyllables are attested in CELEX (Baayen et al. 1993).

(4) *Examples of segmental restrictions on the second syllable: sonorants but not obstruents*

	Monosyllables			*Disyllables*		*Final segment*
a.	free	freer	i.	angry	angrier	[i]
b.	low	lower	j.	mellow	mellower	[oʊ]
c.	full	fuller	k.	gentle	gentler	[l]
d.	green	greener	l.	common	commoner	[n]
e.	loose	looser	m.	*bogus*	*boguser*	[s]
f.	sick	sicker	n.	*frantic*	*franticker*	[k]
g.	brave	braver	o.	*active*	*activer*	[v]
h.	fit	fitter	p.	*licit*	*liciter*	[t]

For many of the longer adjectives, the phonological restrictions are confounded with morphological complexity. There are very few polysyllabic adjectives that end in obstruents and do not contain the suffixes *-ic*, *-ive*, *-ent*, *-ous*, and so on. Most of the polysyllabic adjectives that end in [n] contain the suffixes *-an* or *-ine* (*Martian, bovine*) or the participial *-en* (*broken*). Thus, the data are consistent with the conclusion that comparative formation is conditioned not by the phonological factors but by the last morpheme in the stem. Alternatively, Mondorf (2009) and Hilpert (2008) suggest that morphological complexity itself could be responsible for blocking *-er* suffixation. For the adjectives that end in *-y* or *-ly*, it has been suggested that the morphological affiliation of [i] plays a role in the restriction: final [li] increases the likelihood of a *more Adj* form being used (Lindquist 2000). We do not explore the issue of morphological complexity further here.

For monosyllabic adjectives, the segmental restrictions can be hard to pin down because there are so many exceptions, and the overall number of adjectives that end in a given segment is in some cases rather low. For example, there are only three adjectives in English that end in [ð]: *smooth, lithe, blithe*. Of these, *smoother* is the only attested suffixed comparative in CELEX. The other two do occur in the Google Ngrams corpus (Michel et al. 2011, the "English 2009" subcorpus), but the English speakers we have informally surveyed find *smoother* to be much more acceptable than *lither* or *blither*. Even if all three of these adjectives occured with *-er* frequently, it would not necessarily mean that the suffix is productive on [ð]-final monosyllables, since productivity usually requires more than three examples of the same type (Albright and Hayes 2003, Bybee 1995, 2001, Gouskova and Becker 2013, and others). Consider another example: of the 10 adjectives that end in [ʧ] in CELEX (*arch, butch, Dutch, French, kitch, rich, Scotch, staunch, last-ditch, top-notch*), only *rich* is listed as having an *-er* form. This takes us to the next batch of factors, which we term "lexical heterogeneity".

1.2.3 Lexical heterogeneity

Even when an adjective meets all the the phonological restrictions on size and segmental content, it will not necessarily combine with the *-er* suffix. Comparative formation depends

on semantic properties of the adjective (Matushansky 2002, Graziano-King and Cairns 2005, Boyd 2007, Hilpert 2008): some adjectives are more scalar than others (e.g., compare *pleasant* and *pregnant*). There are various diagnostics and measures of scalarity. Hilpert uses occurrence with *more* as a measure: the more often an adjective occurs with *more* or *-er*, the more gradable it is. We are interested in phonological factors, so we think the best way to control for scalarity is to use nonce words, which come with no semantics at all.

Differences go beyond semantics. British English texts have more *-er* comparatives than American ones, for example (Mondorf 2009). Such dialect differences are a mystery if only syntax, phonology, and semantics are at play, since the grammars of these dialects are similar. There are also lexical idiosyncrasies: there is very little semantic or stylistic difference between *ill* and *sick*, and yet we find *sicker* but not *iller* (thanks to Larry Hyman, p.c. for this example).

Differences in comparative formation are sometimes said to be associated with frequency (Mondorf 2009, inter alia). For example*, stupid* is considerably more frequent than *vapid* and *rapid*, and *stupider* is attested in the Google Ngrams corpus but *vapider* is not; *rapider* is just barely attested (for all of these, forms with "more" are robustly attested, suggesting that the issue is not one of semantic gradability). Hilpert's (2008) carefully constructed corpus study of English comparatives does not find a correlation between token frequency of the base adjective and *-er* suffixation, however.

Furthermore, usage of the comparative form can change over time for some adjectives. The adjective *pleasant* occurs with *more* at a fairly constant rate in the Google Ngrams corpus, but the use of *pleasanter* fluctuates dramatically over time: it peaks around the year 1880 and has been on decline since then. Compare *stupider*, which has been on a relatively steady rise over the same time period (Fig. 1.2.3). If only phonological and semantic factors mattered, the adjectives should not fluctuate over time: they are both obstruent-final, and so they both violate the generalizations about second syllable consonants. The changes in usage point instead to fluctuating lexical specifications for combining with the comparative.[2]

There is also lexical heterogeneity among suffixes. English has at least two other suffixes homophonous with the comparative *-er*: the nominalizer *-er* found in *nominaliz-er, work-er, hatt-er*, and the demonym *-er*, found in *New York-er*, *London-er*, *Newfoundland-er*. The three *-er* suffixes differ not only in syntactic but also phonological selection. We discuss both suffixes in §2.3.3; here we just note that neither the nominalizer nor the demonym restrict the size of the base to mono- and disyllabicity, and the demonym selects for different segmental content than the comparative does.

To sum up, if a theory aims to explain how phonological factors affect morphological rules such as comparative *er*-affixation, the theory needs to capture phonological factors such as syllable count, stress location, and the segmental restrictions on the second syllable, but it also needs to explain why not all adjectives that meet the phonological criteria are attested with the suffix—and conversely, why some adjectives do combine with the suffix

[2]Perhaps the most striking example of a gradation affix on a phonologically deviant base is *winningest*, which has been on a steady rise in sports commentary over the past 40 years. We do not discuss superlatives here, but it bears pointing out that many of the cases that are marginal with *-er* improve when suffixed with *-est* instead.

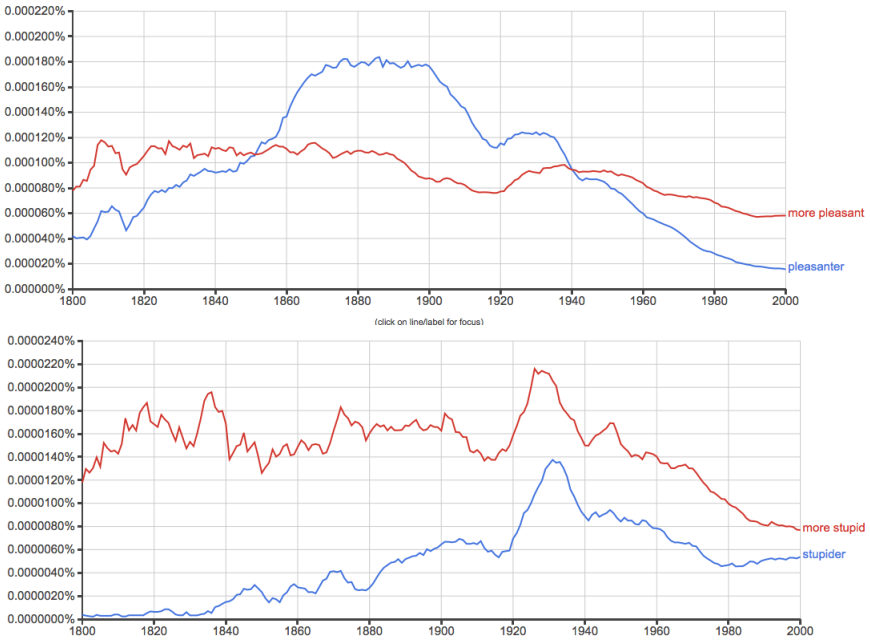


Figure 1: Lexical differences between adjectives, tracked over time

despite failing the phonological criteria. An adequate theory should make predictions for how humans generalize these affixes to novel words, where stylistic, semantic and lexical factors are controlled for. We turn to our theory next.

2. Sublexical Phonotactic Grammars

2.1 The theory

The theory of phonological selection and generalization that we adopt is Sublexical Phonology (Gouskova et al. 2015, Becker and Gouskova 2016, Becker and Allen 2015). This theory holds that learning lexically specific morphological and phonological rules involves separating the lexicon[3] into *sublexicons.* Phonological generalizations about the application of such rules are encoded in part as phonotactic grammars learned over sublexicons. The sublexical phonotactic grammars are similar to language-wide phonotactic grammars. While language-wide phonotactic grammars encode static knowledge about the types of words that languages have, sublexical phonotactic grammars encode static knowledge about the items that undergo a particular rule. Language-wide phonotactic grammars can be used to decide whether a new word sounds native; sublexical phonotactic grammars can be used to decide whether a novel item can undergo a particular rule. We assume that phonotactic grammars assign probabilities to words and that they are learned by feature-based

[3]The idea that the lexicon is separated into strata appears in some form or another in most theories of lexically specific phonological and morphological rules (Mohanan 1982, Kiparsky 1982, Inkelas and Orgun 1995, Ito and Mester 1995, Benua 1997, Pater 2000, Zuraw 2000, Becker 2009, Booij 2011).

generalization over attested sequences; un- and under-attested sequences are penalized by weighted constraints (Vitevitch and Luce 1999, Hayes and Wilson 2008, Albright 2009a, Adriaans and Kager 2010, inter alia).

Here is how this theory can be applied to comparative formation. Learners notice that the morphological alternation *Adj*∼*Adj-er* is attested for some adjectives but not others. They start accumulating a sublexicon of adjectives that they encounter with the comparative *-er* suffix: {*big, sad, busy, blue, bitter, . . .*}. This set, the *-er*$_{comp}$ sublexicon, is characterized by certain phonotactic generalizations covering stress location, the position and prevalence of unstressed syllables, and restrictions on the content of these syllables, as well as the more familiar phonotactic restrictions against final clusters of consonants and so on. For example, unstressed syllables are comparatively underattested initially compared to finally; there are no sequences of stressed syllables or unstressed ones. The sublexicon is used grammatically in the rule for realizing the comparative: existing adjectives form comparatives through suffixation only if they are listed. The sublexical phonotactic grammar can also be used to assess the likelihood of a nonce word belonging to the *-er* sublexicon. Thus, this can be said to be a theory of analogy: it extends rules to new words using the phonological similarity of these new words to some existing set. Because our phonotactic grammars are constraint-based, the decision of whether to extend the rule is based not on sharing a property (e.g., rhyming with *sad*) but rather on properties that an adjective lacks (e.g., secondary stress, unstressed initial syllables, etc.).

2.1.1 What is the content of sublexicons?

We used the purposefully vague term "items" above to refer to the content of sublexicons. In the context of our wug experiment, "items" is synonymous with "morphemes". But some stems for *-er* are morphologically complex (e.g., *eas-y*, *un-tidy*, etc.). The theory must address the step between learning phonological generalizations over phonological words and encoding them into rules stated over morphemes, which we assume to be the units that the grammar manipulates (Embick and Marantz 2008, Gouskova 2012, Gouskova and Linzen 2015, Embick 2015). One possible route to this involves morphological decomposition. In the English comparative sublexicon, a disproportionate number of items end in unstressed [i]. Of the 212 disyllabic adjectives listed in CELEX as having an *-er* form, 183 end in [i]. We assume that the learner notices that these adjectives share a morpheme (the vast majority of them are uncontroversially complex) and formulates a potentiation rule. According to this rule, whenever the Adjective category is expressed as *-y*, the comparative is expressed with *-er*:

(5) Adjective-forming *-y* potentiates *-er*
COMP ↔ [-ɪ] / [-i] $_{\text{Adj}}$ ——

This predicts that [-i]-final adjectives should combine with *-er* even if they exceed two syllables in size, and indeed such comparatives are common in the wild.

This treatment of *-y* reanalyzes the conditioning of *-er* from phonological to morphological. Morphological conditioning of affixation is quite common. In English, nouns are productively formed with *-ness* (as well as other morphs), but adjectives formed with *-able* are nominalized with *-ity*, which is otherwise less productive (Embick and Marantz 2008). In a more general version of this pattern, many languages with inflectional classes have unpredictable gender/inflectional classes for bare roots, but in complex words, the rightmost non-adjunct suffix determines the declensional class of a word (see Wiltschko 2006, Steriopolo 2008, Ott 2011, Gouskova and Linzen 2015). In Russian, the diminutive allomorphs [-ik], [-ok], and [ʧʲik] are for the most part idiosyncratically specified for each root, but the unstressed [-ək] generally appears after diminutive suffixes and [-nʲik] (Gouskova et al. 2015). The generalization about comparison formation of [-i]-final adjectives is subsumed under this pervasive observation about morphology.[4]

2.2 The sublexicon of comparatives: a lexicon study

To expand on the sketch of the sublexical phonotactic grammar account, we want to objectively identify the phonological properties that characterize adjectives that combine with the *-er* comparative suffix in English.

2.2.1 Phonotactics of English comparative adjectives

We extracted all of the adjectives from CELEX (Baayen et al. 1993), and then we searched CELEX for adjectives that appeared with the comparative suffix *-er*. We got American English pronunciations of these adjectives from the Carnegie Mellon University pronunciation dictionary; if an adjective did not appear in the CMU dictionary, it was not used in the lexicon study.[5] This left us with 5345 English adjectives, 400 of which appeared with *-er*.

We trained the UCLA Phonotactic Learner (Hayes and Wilson 2008) on all adjectives and on the 400 comparatives. The learner requires at least 3000 words for training, so we multiplied the list of 400 bases to reach that minimum. We supplied the learner with a list of features marking primary and secondary stress and all of the segmental contrasts of English, based mostly on Hayes (2009) (all of the materials are available at http://gouskova.com/gouskova-and-ahn). The learner was given a default segmental tier and a Vowel tier, which sees the stress features that we specified only on syllabic segments, i.e., [±stress], [±mainstress], and word boundaries. The feature [−mainstress] groups together vocoids that are unstressed or have secondary stress, and it figures prominently in the constraint set that the learner induced. The learner was set to generate 30 constraints, and to consider at most trigrams (i.e., a sequence of three segments or word boundaries).

[4]Some [i]-final adjectives do not appear to be synchronically derived, e.g., *busy, savvy, holy*, *merry*. We would argue that even these words consist of a cranberry root and an affix. This predicts that [-i]-final adjectives longer than two syllables should take comparative *-er*, and this does appear to be the case for many speakers. We do not test this prediction explicitly in our experiment.

[5]We used the CMU dictionary because the participants in our experiment (described in §3) speak American English.

The grammar generated by the learner includes the constraints in Table 1 in the order of weight (=importance). The learner induced several constraints that derive the generalization that English comparative bases tend to consist of a single moraic trochee: the highest-weighted constraint is *CLASH (*[+stress][+stress]), and there are several constraints on secondary stress and positional stress lapses. The learner also induced several constraints whose joint action enforces restrictions on the segmental content of the second syllable, such as *[−tense]# and *[−mainstress,+high][−sonorant]. The *[−tense]# constraint captures in part the observation that [i] is overrepresented word-finally among disyllables in the sublexicon of comparative bases. The constraint *[−mainstress,+high][−sonorant] is a more specific version of a prohibition against obstruent codas in unstressed position, as in "vápid".

2.2.2 Phonotactics of English adjectives

For comparison, consider the constraints the learner induced after training on all English adjectives. *CLASH is not in the set—there is a constraint against a sequence of two stressed segments on the default tier, but since stress is only specified for syllabic segments, it is really a constraint against two stressed vowels in hiatus (as in [dùɛ́t]). There are constraints on stress position, e.g., *[+stress][+mainstress]# on the vowel tier penalizes words that have the profile [...σ̀σ́], and *#[-stress][-mainstress] penalizes words that have the profile [σσ̀...], but these are more specific than what we see in the comparative base grammar.

2.2.3 The status of rhotics in different dialects

Note what is missing from the comparative sublexicon phonotactic grammar in Table 1. When trained on comparative adjective bases, the learner did not induce constraints that we might expect based on the generalizations in Mondorf (2009): there is no penalty for final consonant clusters or for final [ɹ] (see also Hilpert 2008). When we supplied the learner with handmade constraints penalizing final clusters and [ɹ] and gave them an initial weight of 6, it weighted them very low—they were in the bottom four. When we used this manually rigged grammar to analyze the results of the nonce study we describe in §3, it did not do as well as the grammar that the learner generated on its own. This suggests that final rhotics are not underattested in the training set, at least given the thresholds assumed by this learner.

It is unclear whether the CELEX model of the Adj-er sublexicon reflects the phonological status of [ɹ] in comparative formation in American English. Recall that the CELEX database is compiled on the basis of British English usage, and Mondorf's generalizations are based on counts in British newspapers. Many British dialects are non-rhotic; in such dialects, the suffix *-er* is pronounced as a schwa, whereas in rhotic ones, it is realized as a flap or the highly marked English liquid [ɹ]. Avoidance of [ɹ] sequences might discourage attachment to r-final adjectives in rhotic dialects, which we found to be a strong

Constraint	Tier	Weight
*[+stress][+stress]	Vowel	6.443
*[−mainstress,+tense][seg]	default	5.38
*[−mainstress,−high,−back]	default	4.871
*[−sonorant][+voice]	default	4.82
*[+high,−syllabic][−syllabic]	default	4.772
*[−tense]#	default	4.649
*[−continuant][+nasal]	default	4.514
*[−mainstress,+high][−sonorant]	default	4.374
*[−mainstress,+low]	default	4.309
*[−mainstress,+round,+high]	default	4.267
*[+sonorant,-syllabic][−consonantal,+sonorant,−syllabic]	default	4.252
*#[−mainstress]#	Vowel	4.244
*[+consonantal,+sonorant][+nasal]	default	4.194
*[+voice][−sonorant]	default	4.191
*[+continuant,+voice,+CORONAL][−syllabic]	default	4.156
*[−sonorant,+CORONAL][+continuant]	default	4.107
*[−syllabic][+continuant,+LABIAL]	default	4.096
*[+consonantal][−voice,−anterior]	default	4.085
*[+continuant,+voice,−anterior]	default	4.054
*[+consonantal,−anterior][+consonantal]	default	3.976
*[+LABIAL][+continuant,−voice]	default	3.91
*[−round][−low]	default	3.892
*##	Vowel	3.846
*#[+round,+high,+syllabic]	default	3.846
*[+diphthong,−mainstress]	default	3.738
*[seg][seg][+stress]	Vowel	3.432
*[−mainstress][−mainstress]	Vowel	3.385
*[−round][+syllabic]	default	2.085
*[+syllabic][+stress]	default	2.036
*[−mainstress,+stress]	Vowel	1.667

Table 1: Constraints induced by the UCLAPL when trained on English comparative adjective bases

Constraint	Tier	Weight
*#[−mainstress]#	Vowel	5.631
*[−tense][+syllabic]	default	5.126
*[−tense]#	default	5.122
*[+low][+syllabic]	default	4.602
*[−voice][+continuant,+voice]	default	4.556
*[+low]#	default	4.548
*#[−diphthong,+round,+high]	default	4.527
*[+diphthong,−mainstress]#	default	4.504
*[+round,+syllabic][+round]	default	4.463
*[+high,−syllabic]#	default	4.424
*##	Vowel	4.421
*[−back,+syllabic][+high,+tense]	default	4.152
*[+voice,+CORONAL][+consonantal,−anterior]	default	4.065
*[−stress,+round,-back]	default	4.021
*[+consonantal,-anterior][+continuant,+CORONAL]	default	4.005
*[+stress][-stress,−high,−back]	default	3.994
*[+stress][+stress]	default	3.806
*[−mainstress,+stress,+back][−mainstress,+tense]	default	3.779
*[−mainstress,+stress][−mainstress,+stress]	Vowel	3.652
*[−mainstress,−high,−back][−mainstress]	default	3.544
*[+stress][-stress,+low]	default	3.543
*[+stress][+mainstress]#	Vowel	3.107
*#[−stress][−mainstress]	Vowel	3.094
*[−mainstress,+stress][seg][−mainstress,+stress]	Vowel	2.903
*[+mainstress][seg][+mainstress]	Vowel	2.643
*[+mainstress][+mainstress][seg]	Vowel	2.53
*[seg][+stress][−mainstress,+stress]	Vowel	2.319
*[seg][+stress][+mainstress]	Vowel	0.573
*[−stress][−mainstress,+stress][−mainstress,+stress]	Vowel	0.1
*[+stress][+stress][+stress]	Vowel	0

Table 2: Constraints induced by the UCLAPL when trained on all English adjectives

effect in our study (see §3). Mondorf speculates that the difference between American and British rates of *more Adj* and *Adj-er* comparatives are due to stylistic preferences in syntactic/morphological complexity. One could test the consequences of such differences by comparing *er*-suffixation rates between non-rhotic and rhotic dialects in America, and non-rhotic and rhotic dialects in the UK. We have not done this, so it remains an open question. We accept that there are limitations to our ability to model speakers' internal sublexicons with corpus-based models such as our CELEX set.

2.3 Other approaches to phonological selection

Before moving on to our experiment, we want to cover some alternative approaches to the selectional restrictions of comparative *-er,* to set up the predictions they make for people's behavior.

2.3.1 Subcategorization frames

Subcategorization frames are the oldest and most popular approach to phonological selection. These are morpheme-specific rules that encode the phonological or morphosyntactic context of each affix/allomorph (Lieber 1980, Paster 2006, Bye 2007, Embick 2010, inter alia). Thus, McCarthy and Prince (1990:236) characterize the affixes as subcategorizing for minimal phonological words consisting of a single foot—in English, a moraic trochee (single heavy syllable or ˈH or a pair of lights with initial stress, ˈLL). The subcategorization frame would then refer to the "minimal word" in its context, as in (6):

(6) *A subcategorization frame approach to English comparatives:*
COMP $\leftrightarrow$ -ɪ / [ ]$_{\text{MinWd}}$ ——

This kind of approach is often implicitly assumed but rarely formalized in the literature on English comparatives. One difficulty with it is pinning down exactly what unit is subcategorized for. Different affixes in the same language can subcategorize for differently sized units (see Downing 2006 for related discussion). In English, the suffix *-en* (*whiten*, *stiffen*, etc.) selects strictly for heavy monosyllables—which also happens to be the size of the minimal word in the language (Siegel 1974 et seq.). The rule in (6) does not explicitly explain why the comparative tolerates mono- and disyllables in its bases but the verbalizing suffix *-en* only selects monosyllables. The context would have to be [( )$_{\text{Ft}}$]$_{\text{Pwd}}$ for comparatives and [($\sigma_{\mu\mu}$)$_{\text{Ft}}$]$_{\text{Pwd}}$ for the verbalizing *-en*.It turns out to be difficult to formalize the slightly different prosodic conditions of the affixes into elegant subcategorization frames.

Another problem with this approach is that it fails to capture the restrictions on the content of the second syllable in the base. The subcategorization frame account does not have an explanation for why the content of the coda should matter only for disyllabic bases; correspondingly, while the segmental restrictions on the second syllable are often noted, they are never formally accounted for. It is even harder to accommodate lexical differences between foot-sized adjectives that do and do not combine with the affix (*stupid* vs. *vapid*,

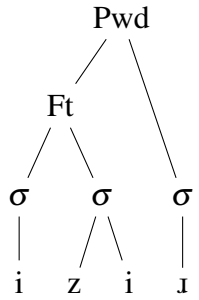
ill vs. *sick*, etc.). If the motivation for formulating the subcategorization frame rule is to capture some productive aspect of the affix's distribution, then it is a mystery why so many adjectives that meet the frame's structural description nonetheless fail to combine with the affix.

2.3.2 Generalized Alignment

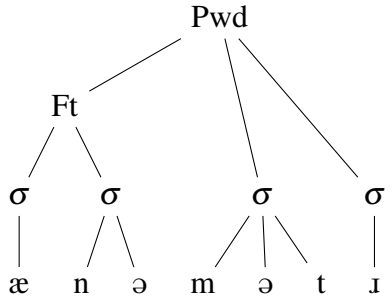
Phonological conditions on affixation are sometimes formalized in terms of Generalized Alignment constraints (McCarthy and Prince 1993). Alignment captures a similar intuition to subcategorization frames: the affix wants to be near some phonological unit. In the case of Alignment, the requirement is formalized as coincidence of edges, which presents problems when the material that prevents affixation appears on the opposite side of the stem from the affix. Suppose the requirement is that *-er* appear at the right edge of the main stress foot, ALIGN(L, -ɪ, R, HdFt). This correctly distinguishes the good *easier* (as in 7a) from the bad *ánimater* (as in 7b), but it cannot distinguish *easier* from the degraded *moroser* (as in 7c). The unstressed syllable [mə] on the left side of the word is invisible to alignment. And just like the subcategorization frame analysis, the alignment analysis requires some additional mechanisms in order to capture the segmental and syllabic factors affecting the distribution of *-er*.

(7) A challenge for an Alignment account of the monopod requirement

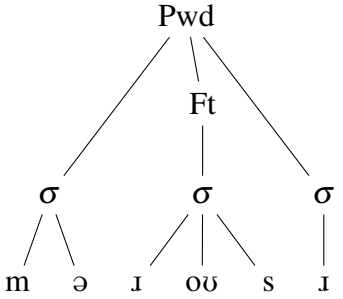
a. *easier: -er is aligned to a foot*



b. **animater: -er is not aligned to a foot, correctly ruled out*



c. **moroser: -er is aligned to a foot (syll on left invisible to alignment)*



The alignment analysis is also very sensitive to syllabification. The trees in (7) assume that *-er* is syllabified alone, but alignment will be violated if it shares the syllable with material from the preceding stem. The syllabification in words like *wet.ter* and *nic.er* is controversial (Kahn 1976, Turk 1994, Durvasula et al. 2013), but some analyses of the [g]∼Ø alternation, as in *longer* [lɑŋgɹ] vs. "longing" [lɑŋɪŋ] rely on the assumption that [g] is syllabified into an onset in *longer* (see Borowsky 1986, a.o.). For this account to explain why [g] is not deleting in *lon.g-er*, one would need to assume that [g] is syllabified with *-er*, which means that the form also violates the alignment constraint. Once again, while Alignment captures part of the intuition about prosodic shapes in *-er*-affixed words, it does not succeed in explaining details.

2.3.3 Emergence of the Unmarked

The Emergence of the Unmarked formalizes the intuition that general phonological constraints can determine the distribution of affixes without having a strong effect elsewhere in the language (McCarthy and Prince 1994, Mascaró 1996).[6] Take a simple example: English syllables are not required to have onsets and are not forbidden from having codas, so the constraints ONSET and NOCODA are dominated by faithfulness. The constraints ONSET and NOCODA emerge in selecting between the two suppletive allomorphs of the indefinite article in English: ONSET prefers *an apple* to **a apple*, and NOCODA prefers *a pin* to **an pin.* Both of the surface forms violate the markedness constraints, but the constraints still have an effect because faithfulness does not matter in selecting between two suppletive allomorphs.

One of the appealing features of this approach is that it can accommodate the effects of multiple constraints, especially if it is implemented in a weighted constraint framework that allows for ganging effects. This is an argument in favor of a TETU-based approach as opposed to Subcategorization Frames, which have difficulty capturing complex multifactorial interactions. On the other hand, markedness constraints often conflict with each

[6]This approach is most often applied to suppletive allomorphy, and the attendant debate as to whether is output-optimizing or not is extensive (Raffelsiefen 1999, Rubach and Booij 2001, Mascaró 2007, Embick 2010, Bobaljik 2000, Wolf 2015, Nevins 2011, Wolf 2013, Gouskova et al. 2015, Bennett 2016). We are not dealing with an obviously suppletive case here, so another potentially relevant debate is whether ineffability and paradigm gaps are conditioned phonologically or lexically (Halle 1973, Hetzron 1975, Orgun and Sprouse 1999, Wolf and McCarthy 2010, Raffelsiefen 2004, Pertsova 2005, Daland et al. 2007).

other, so the ranking needs to be consistent with the phonology of the language. We will argue that this does not hold for the TETU account of English comparatives.

2.3.4 A TETU analysis of English comparatives

What assumptions would be required in order to extend this approach to English comparatives? Comparative formation in English does not involve suppletive allomorphy in an obvious way, so extending this approach to comparative and superlative formation requires assuming either that (i) the *Adj-er* and *more Adj* forms compete in the same derivation, or (ii) there are separate derivations for the *Adj-er* and the *more Adj* forms, but one of them maps to a null output (i.e., results in a gap) in cases where the input for adjectives that normally do combine with the suffix, such as *smarter (*vs. *more smart).* In either case, some phonological constraints must disfavor both *intelligenter* and *more smart*.

Which constraints do the work? There are no formally developed Emergence of the Unmarked accounts of the comparative alternation that we know of, but some detailed descriptive studies suggest a few possibilities. Mondorf (2009) observes that initially stressed monosyllables have stress clashes in the more Adj form, *móre smárt*, and therefore tend to surface as *Adj-er* more often. We could further recruit *LAPSE to rule out *intélligenter* in favor of *móre intélligent* (see Elenbaas and Kager 1999 inter alia). Additional factors include an OCP-like constraint against adjacent [ɪ] sequences (as in *sore/sorer*) and consonant clusters (as in *honest/honestest*, *moist/moistest*)—Mondorf finds that in corpora of newspaper English, such *Adj-er* forms are less common than their *most Adj* forms. Some of these effects are confirmed in the systematic corpus study by Hilpert (2008), as well.

We implement this analysis in (8), deriving the correct outcomes for *apt, intelligent, morose, silly, black*, and *sore*. The constraints are *r-r 'assign a violation mark for any [ɪ] preceded by an adjacent [ɪ]', *TT-r 'assign a violation mark for an [ɪ] preceded by an obstruent cluster', and the familiar PARSE-σ, *CLASH and *LAPSE (McCarthy and Prince 1993, Elenbaas and Kager 1999). The ranking of some of these constraints is underdetermined in English, but at least PARSE-σ must be ranked above *CLASH based on stress patterns in longer words with multiple heavy syllables (Pater 2000).

(8) An analysis of comparative formation in the Emergence of the Unmarked

	/Adj {-ɹ, moɹ}/	*r-r	*TT-r	PARSE-σ	*CLASH	*LAPSE
a.	→ moɹ æpt (σ́)(σ́)				*	
b.	æp.tɹ (σ́)σ		*W	*W	L	
c.	→ moɹ ɪn.tɛ.lə.ʤənt (σ̀)σ(σ́σ)σ			**		*
d.	ɪn.tɛ.lə.ʤən.tɹ σ(σ́σ)σσ		*W	***W		**W
e.	→ moɹ mə.roʊs (σ̀)σ(σ́)			*		
f.	mə.ɹoʊ.sɹ σ(σ́)σ			**W		
g.	→ sɪ.li.ɹ (σ́σ)σ					*
h.	moɹ sɪ.li (σ́)(σ́σ)				*W	L
i.	→blæ.kɹ (σ́σ)					
j.	moɹ blæk (σ́)(σ́)				*W	
m.	→moɹ soɹ (σ́)(σ́)	*		*		
n.	soɹ.ɹ (σ́)σ	W		W	*L	

2.3.5 Challenge for TETU: affix-specificity of restrictions

One challenge for this account is explaining lexical differences between bases. We have already discussed phonologically similar adjectives that do and do not combine with *-er*: *taller/*iller*, *stupider/*vapider*, *graver/*mauver.* Either the *Adj-er* or the *more Adj* norms have to be treated as exceptions. The most attractive aspect of this approach is that it uses independently motivated constraints without an appeal to lexical stipulations about affixal subcategorization, but the facts of English require lexical listing of adjectives that combine with *-er*.

Another issue is that these constraints should act the same way on other *-er* suffixes, and they do not. The problem in a nutshell is that English has at least two more suffixes that are homophonous with *-er*, and these suffixes either exhibit no phonological selectional restrictions or impose restrictions that contradict those of comparative *-er*.

Nominalizing *-er* Let us start with the affix that does not impose phonological selectional restrictions: nominalizing *-er*. English bases for *-er* nominalizations routinely exceed two syllables in size *(malinger/malingerer, experiment/experimenter),* and the disyllables do not need to have initial stress (*redeem/redeemer*, *attack/attacker*). If *LAPSE acts on the comparative *-er* in **beautifuler* (σ́σσσ), why does it not likewise rule out *experimenter (σσ́σσσ)?* One objection to this argument[7] is that comparative derivation is paradigmatic: i.e., if the *Adj-er* form does not exist, the *more Adj* form is usually possible. On the other hand, in the nominalizer *-er* derivation, there is no "alternative candidate". The counterargument is that plenty of English adjectives lack comparative forms altogether, and some

[7]Thanks to Elliott Moreton and Jochen Trommer, p.c., for discussing this alternative with us.

have both a "more" and an *--er* form, as documented in the rich literature on English comparatives cited earlier. As for nominalizers, English does have alternatives to *-er*: *-ist, -ent*, and so on (Rodríguez and Quintero 2002).

(9) Other nominalizer suffixes in English: allomorphs of *-er*?

a.	work	worker	e.	machine	machinist
b.	hat	hatter	f.	study	student
c.	malinger	malingerer	g.	cook	cook
d.	diary	diarist	h.	plow	plowman

Whether this is a case of suppletive allomorphy or synonymy is a matter of debate, but we believe this is a case of suppletion. Nominalizing suffix allomorphy appears to be not phonologically conditioned but rather lexically determined by the preceding morpheme. Some of the *-er* nouns have hallmarks of so-called root nominalizations. In such words, the affix assigns a category to a root that has none (Embick and Marantz 2008 and others). The free bases of *-er* nominalizations, where they exist,[8] vary in category from verbs (*work*, *entertain*) to nouns (*hat*, *roof*), and the meaning of the resulting derived noun is not straightforwardly compositional: a *worker* is someone who works, but a *hatter* is someone who makes hats. This is a property shared by all the forms in (9): e.g., a *milkman* isn't necessarily someone who milks, but rather delivers milk. Yet another important feature of root nominalizations is lexically conditioned allomorphy: the pronunciation that realizes the noun head {*-er, -ist*, etc.} is idiosyncratically listed for each root; on the other hand, allomorphs realizing the nominalizing head *n* after overt category affixes tend to be more predictable (e.g., after *-ize*, the nominalizer is *-er*, never *-ist*). These facts suggest an analysis whereby bases associated with *-er* and *-ist* are lexically listed.

To assess quantitatively whether the phonotactic restrictions of nominalizing *-er* are stricter than those of adjectival *-er*, we ran a simple learning simulation using the UCLA Phonotactic Learner. We trained phonotactic grammars on adjectives that occur with comparative *-er* and on verbs that occur with nominalizing *-er* in CELEX. We then tested the grammars on two random samples from CELEX: a sample of all English verbs, and a sample of all English words. The samples were of a comparable size to the sublexicon of verb bases for nominalizing -er, around 3500 words. The phonotactic scores assigned by the UCLA Phonotactic learner are not directly comparable between the adjective-trained and the verb-trained grammars, because constraints are weighted on scales determined by different datasets, but we can compare the percentages of all words and all verbs that receive zero violation marks in each grammar (see Table 3). The difference is striking: the chances of any given random English verb passing the grammar trained on "work"-type verbs is 82%; on the other hand, the chance of any given random English verb passing the grammar trained on "big"-type adjectives is only 56%. The differences persist when we broaden the comparison to all English words, which are more phonologically heterogeneous than

[8]Plenty of compound nominalizations are not formed from free-standing bases: *scene-stealer, truck driver*, etc. (Giegerich 2004, McIntyre 2014, inter alia). These examples highlight that there is considerably more to well-formedness in *-er* nominalizations than phonology; argument structure and semantics matter as well.

	Grammar trained on *-er* adj	Gr. trained on *-er* verbs
Random sample of verbs	56%	82%
Random sample of all words	31%	45%

Table 3: Percentages of English verbs and English words that pass the UCLA Phonotactic Learner grammars trained on comparative adjectives (e.g., "big") and verbs that occur with *-er* (e.g., "work") in CELEX.

verbs (Guion et al. 2003, Albright 2008, and others). These words include complex derived nouns such as "industrialization", which have no hope of combining with nominalizing *-er*, so their failure to receive high passing marks in that grammar is not surprising.

These simulations lend some objective support to the impressionistic conclusion that the *-er* that attaches to verbs is not nearly as phonologically selective as comparative *-er*. It favors verbs more than English words overall, but compared to the adjectival *-er*, the nominalizing *-er* is rather phonologically inclusive.

Demonym -er There is of course yet another *-er* affix in English, the demonym[9]:

(10) Demonym *-er* vs. *-an*

New York	New Yorker	Philadelphia	Philadelphian
Long Island	Long Islander	San Francisco	San Franciscan
Pittsburgh	Pittsburgher	Tacoma	Tacoman
Connecticut	Connecticuter	Miami	Miamian
Vermont	Vermonter	Ohio	Ohioan
London	Londoner	Denver	Denveran

It has been suggested that the distribution of the demonym *-er* is phonologically conditioned (O'Grady et al. 2005:pp. 143–144), and that it is in a suppletive relationship with *-an, -ese,* and *-ite* (Gordon 2014). Gordon observes, among other things, that the demonym *-er* does not attach to vowel-final words—which is relevant to our concerns, as this aversion is in striking contrast to the comparative *-er*. The examples above also make it clear that the size and stress restrictions are not the same for the three *-er* suffixes of English. The case of demonyms, then, seems to be similar to other cases of phonologically conditioned suppletion: the bases are lexically listed, and they tend to follow familiar morphological trends (e.g., *-burgh, -town* and *-land* productively potentiate *-er*, presumably because these morphemes are listed in the sublexicon for demonym *-er*). The morphemes that combine

[9] "Connecticuter" is listed in many official manuals as the term for a denizen of Connecticut but apparently does not enjoy widespread use (see Safire 1982, Dickson 1990). The fact that it even occurs to non-linguists as a possibility suggests that it has a non-negligible probability with respect to the grammar of demonym *-er,* although this could also point to a case of ineffability where speakers are uncertain as to how to select between several bad outputs (see Albright 2009b, inter alia).

with the demonym *-er* are characterized by their own phonotactic generalizations, distinct from those of comparative *-er*.

To conclude, there is no reason for the phonological constraints that emerge in comparative formation to not similarly emerge in nominalization or demonyms. To allow TETU to be affix-specific would require indexing constraints to the individual morphemes somehow, but then this would no longer be a theory that leaves selection to the phonological component alone (see Paster 2006, Bye 2007, Wolf 2015, Nevins 2011 for related discussion). In our theory, both the comparative *-er* and the demonym *-er* have a lexically determined distribution; the distribution of nominalizing *-er* might also be lexically determined but with very few phonological generalizations about bases that combine with it. Comparative bases have more in common with each other phonologically—i.e., their sublexicon is more phonotactically restricted—than *-er* nominalizations.

3. Experiment

3.1 Predictions and design considerations

We now turn to an experiment that tests the predictions of the sublexical phonotactic theory of phonological selection. Our theory predicts that a nonce word will be more likely to combine with the comparative *-er* the more phonotactically well-formed it is in the sublexicon of English adjectives that combine with the suffix. In the experiment, we use an array of nonce words that from a spectrum of phonotactically well-formed, intermediate, and ill-formed words compared to the kind of words that the English *-er* adjective sublexicon contains. The nonce adjectives (described in more detail below) range from monosyllabic to trisyllabic, with a variety of stress patterns. This design allows us to test our theory against the simpler alternative suggested by the subcategorization frame account: the only relevant factor is whether the nonce word is a moraic trochee. If other phonological factors matter as well, it would support either our account or the emergence of the unmarked.

To distinguish our theory from the emergence of the unmarked, we manipulate the context of presentation: the base *(wug)* and suffixed form *(wugger)* are presented as adjective-comparative or as verb-nominalizer. We would expect the same wug to be rated higher when it is presented as a bare adjective than when it is presented as a bare verb. English verbs are phonologically different from adjectives and nouns; both stress and syllable structure vary between verbs and adjectives in particular (Kawamoto and Farrar 1990, Kelly 1992, Albright 2008, Smith 2011). Verbs are special in the context of a wug study since wugs may be treated as loanwords, and verbs are less likely to be borrowed than nouns (Haspelmath and Tadmor 2009). Some subtlety is needed in discussing adjective borrowing, since adjectives pattern either with nouns or with verbs depending on the language. In general, theories of syntactic categorization either reserve a special category for adjectives and adverbs or lump them with verbs in some languages and with nouns in others (Baker 2003, Borer 2005, Embick 2015). For us, the crucial observation is that adjectives and verbs are phonologically different, since such differences will have an effect in a rating study. The main difference we are interested in, however, is between the rating of *wugger* when presented as a comparative adjective vs. a deverbal noun. Since the nominalizer *-er* is

considerably more productive than the comparative *-er*, the ratings of comparatives should be lower.

Finally, one might ask why we are doing a rating study as opposed to a forced choice experiment (*more wug* vs. *wugger*). The answer is that the two comparative forms are not in complementary but overlapping distribution (Hilpert 2008, LaFave 2015, Lindsay and Aronoff 2013, Matushansky 2013). Many if not most adjectives can occur in the *more Adj* context, but not all can occur in the *Adj-er* context (Di Sciullo and Williams 1987, Embick 2007, Lindsay and Aronoff 2013). A forced choice task does not allow us to decipher whether an adjective is acceptable in both contexts or in neither. Moreover, existing experimental evidence suggests that people prefer *more Adj* whenever the processing load increases (Boyd 2007), which might be the case in nonce word tasks. Since we are primarily interested in the phonological factors controlling the distribution of adjectival *-er* and not in the broader question of comparative formation in English, the rating paradigm is more appropriate.

3.2 Participants

We recruited 267 people on Amazon's Mechanical Turk (http://mturk.com), paid $1 for their time. IP addresses were limited to the USA, and we asked people to self-eliminate if they were not native English speakers. The average time to complete the experiment was about 10 minutes. People were about evenly divided between the two conditions: 134 saw the items presented as Verbs, and 133 saw them presented as Adjectives. We eliminated 24 outlier participants whose correlation with the group for each base rating was more than 0.8 standard deviation points higher than average. This left 124 Verb raters and 119 Adj raters.

3.3 Procedure

People in the Adjective condition heard an audio recording of a nonce word and saw it written in a sentential context that suggested an adjective. They were asked to rate the nonce adjective on a scale of 1–5. We then played the morpheme with *-er* attached, and showed a sentence where it appeared in an adjectival comparative context.

The very same stimuli were also played to the Verb condition group and shown in verbal contexts and with the nominalizer *-er* suffix. Just as in the Adjective condition, everyone rated both the base and the suffixed form.

3.4 Materials

We used 292 nonce words. The words were created as follows. In the process of learning, the UCLA Phonotactic learner (Hayes and Wilson 2008) generates a list of constraints on segmental bigrams and trigrams, and a list of nonce words known as the "sample salad". The sample salad is generated at random from the segments defined in the feature file,

Please push the "play" button to listen to the adjective:

I know you think I am **chike**. ►

Rate the adjective **chike**:

awful [1] [2] [3] [4] [5] good

Now listen to the comparative form:

But you are even **chiker**. ►

How likely would an English speaker be to say **chiker** in this sentence?

not likely [1] [2] [3] [4] [5] pretty likely

Figure 2: Example trial in the Adjective condition

Please push the "play" button to listen to the verb:

Flo recently learned how to **chike**. ►

Rate the verb **chike**:

awful [1] [2] [3] [4] [5] good

Now listen to the noun:

She's an amateur **chiker**. ►

How likely would an English speaker be to call someone who **chikes** a **chiker**?

not likely [1] [2] [3] [4] [5] pretty likely

Figure 3: Example trial in the Verb condition

and it covers the phonotactic probability distribution in the training set. Our stimuli come from the sample salad generated by training on the entire adjective sublexicon; we chose a subset that represented good, medium and poor scores in the comparative sublexicon. We ensured that all of the stimuli were phonotactically well-formed given the grammar that the UCLAPL came up with after it was trained on all adjectives of English. A full list of the stimuli is given in the Appendix, which also contains the 60 pairs of frame sentences we used. The frames were paired with items randomly.

We recorded each nonce word with and without the *-er* suffix, pronounced by a phonetician who was native speaker of a rhotic dialect of American English[10]. The frames were presented only orthographically, but the target bases and derived forms were both played and shown in orthography. Since the stimuli were generated as pronunciations, we made up the orthographies that seemed most likely to match those pronunciations (e.g., [ɑdðɛ́ɪ] as "audthey", [fùdɛ́dəw] as "fudaddow").

To keep the experiment to a manageable length, we presented each person with a structured subset of the 292 bases, with 20 randomly chosen monosyllables, 20 disyllables, and 20 trisyllabic words, for a total 60 base-derivative ratings per person. Each person rated each nonce word once in the course of the experiment. Each base-derivative pair was rated by at least 14 different participants (mean=50, sd=33).

[10]Special thanks to Lisa Davidson for being our talker.

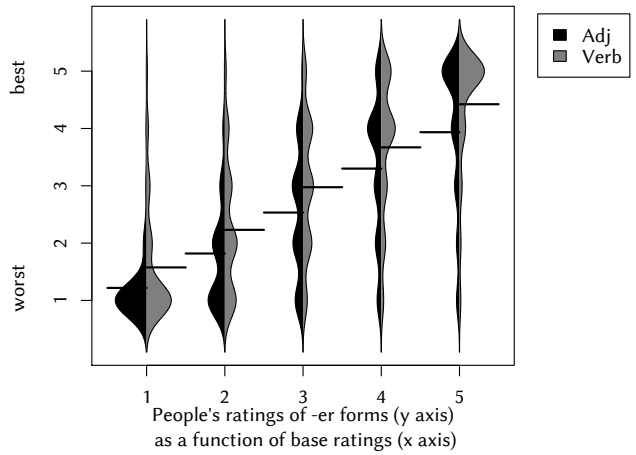


Figure 4: Rating of bare form predicts rating of suffixed form

3.5 Results

The metalinguistic nature of rating tasks sometimes results in people working out a strategy in uncertain cases. In this experiment, the rating of the suffixed form depended more than anything on the rating of the bare form, both for verb and adjective conditions. People most often gave the same or similar rating to a suffixed form as they did to the bare form. Thus, in analyzing the ratings of suffixed forms, we have to take into account the rating of the bare form, and any factor that contributes above and beyond the bare form rating can be interpreted as linguistically interesting.

As expected, nonce words were rated lower when they are presented as verbs (e.g., "to zice", mean rating 2.85) than when they are presented as adjectives (e.g., "very zice", mean rating 3.15). When the words were affixed with *-er*, however, the ratings flipped: deverbal nominalizations ("an amateur zicer") were rated higher (mean rating 2.86) than adjectival comparatives ("much zicer", mean rating 2.66). Note, however, that verb bases ("to zice") were rated about the same as nominalizations ("an amateur zicer"); see Fig. 5. This is a beanplot—a vertical density plot with a horizontal line marking the mean.

To examine the more fine-grained phonological patterns, consider how the ratings break down by the final segment of the nonce word when it is presented as Adj or Verb vs. Adj-er or Verb-er. The ratings are summarized in the beanplots in Fig. 6. As shown in the left-hand plot, the ratings of "Adj" nonce words were consistently higher than "Verb" regardless of the quality of the final segment. On the other hand, the ratings of "Verb-er" nonce words were higher than "Adj-er" when the nonce word stem ended in a non-low vowel (e.g., "poylfee-r" [pójlfiɹ]), nasal (e.g., "chem-er" [ʧɛmɹ], or obstruent ("wurch-er" [wɹʧɹ]), and

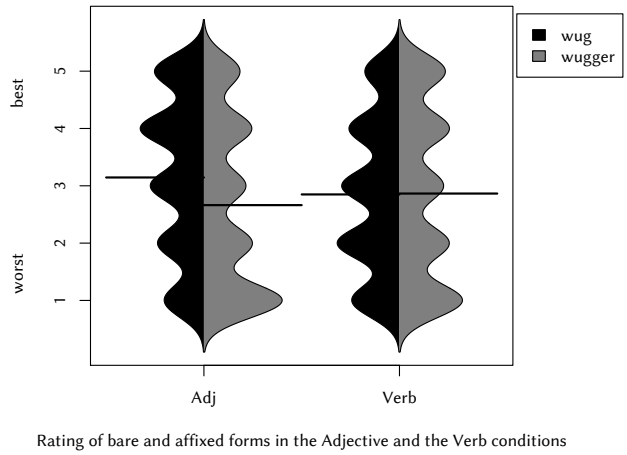


Figure 5: Effect of affixation on ratings, adjective vs. verb presentation

both Adj-er and Verb-er nonce words were rated lower when they ended in [ɑ] or [ɹ] or [l] (e.g., "wunpaw-er" [wunpɑɹ], "nerr-er" [nɹɹ], "joyl-er" [ʤojlɹ]).

The interaction between ratings and context ("Adj"≻ "Verb", but "Verb-er"≻ "Adj-er") persists across stimuli regardless of stress location and type (see fig. 7). People liked monosyllables such as [gɛf] best; disyllables such as [ɪpi] were in the middle (with trochees and iambs about equal), and trisyllables such as [fu'dædəw] were rated the lowest.

In general, people rated words with secondary stress (e.g., "zerpoo" [ˈzɹˌpuw]) lower than words without secondary stress (e.g., [ˈɪpi]). This was true even when monosyllables were excluded from the data (this subset is plotted in Fig. 8). Ratings for stimuli with secondary stress rose slightly when items were presented as adjectives compared to verbs, and this reversed when the *-er* suffixes were added.

Finally, Fig. 9 shows the variation between items' mean ratings in different conditions. The plot on the left shows the ratings of nonce words presented as Adj-er as a function of the phonotactic violations on the x-axis. The plot on the right shows the ratings for the same nonce words presented as Verb-er. The blue line is a fitted linear regression line, with the gray area showing the 95% confidence region. The slope of the line is steeper in the left plot than in the right plot, which reflects the closer correlation between people's ratings and violations in the Adjective condition compared to the Verb condition (a simple linear model for Adj, R^2=0.1799, F(1, 7138)=1565, p¡0.0001; for Verb, R^2=0.1506, F(1,7438)=1319, p¡0.0001). We now turn to a more complete analysis of the results.

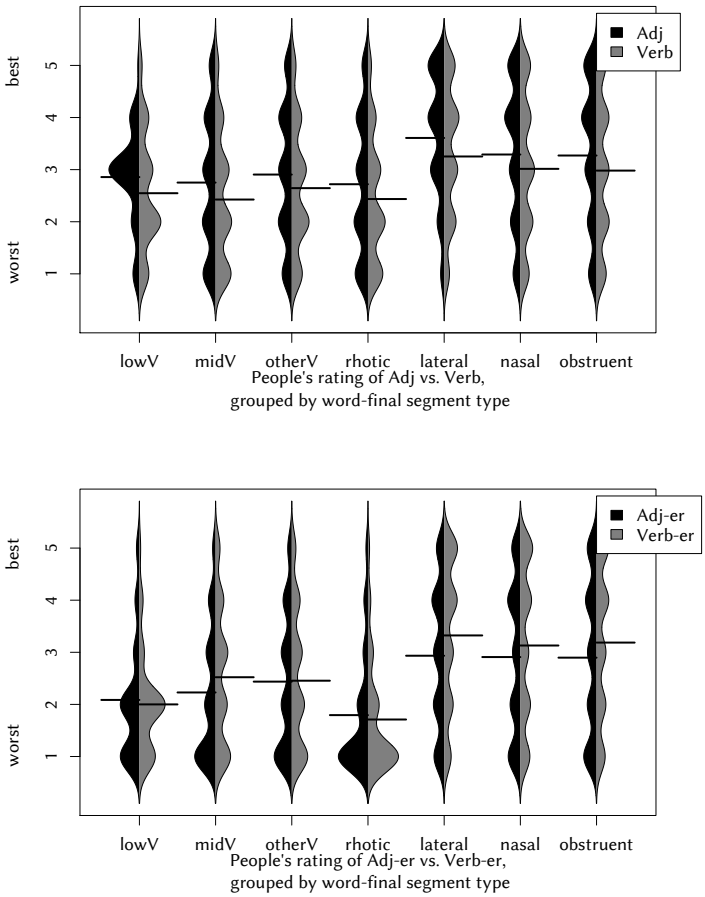


Figure 6: Ratings of nonce words presented as verbs vs. adjectives: stem-final segment

3.6 Statistical analysis of ratings

We analyzed people's ratings of bare forms first, and then we analyzed suffixed forms. The model of bare form ratings is a linear hierarchical model fitted using the *lme4* package (version 1-1.17, Bates and Maechler 2009) in R (R Development Core Team 2013). The p-values were estimated from *t*. The model included a random by-participant intercept and slopes for syllable count, trial number, and base violation score; there were by-item and by-frame random intercepts, and a random by-item slope for experiment name and trial number. This is the most complex hierarchical model allowed by the structure of the data (see Barr et al. 2013). We report two measures of collinearity: variance inflation factor and maximum correlations between predictors. Residualization was not used (see Belsley et al. 2004, Wurm and Fisicaro 2014). We verified that residuals were normally distributed by visually inspecting a residuals plot for all the models reported here.

As can be seen from the coefficient estimates in Table 4, people give slightly higher ratings as the experiment progresses—the effect of trial number approaches significance

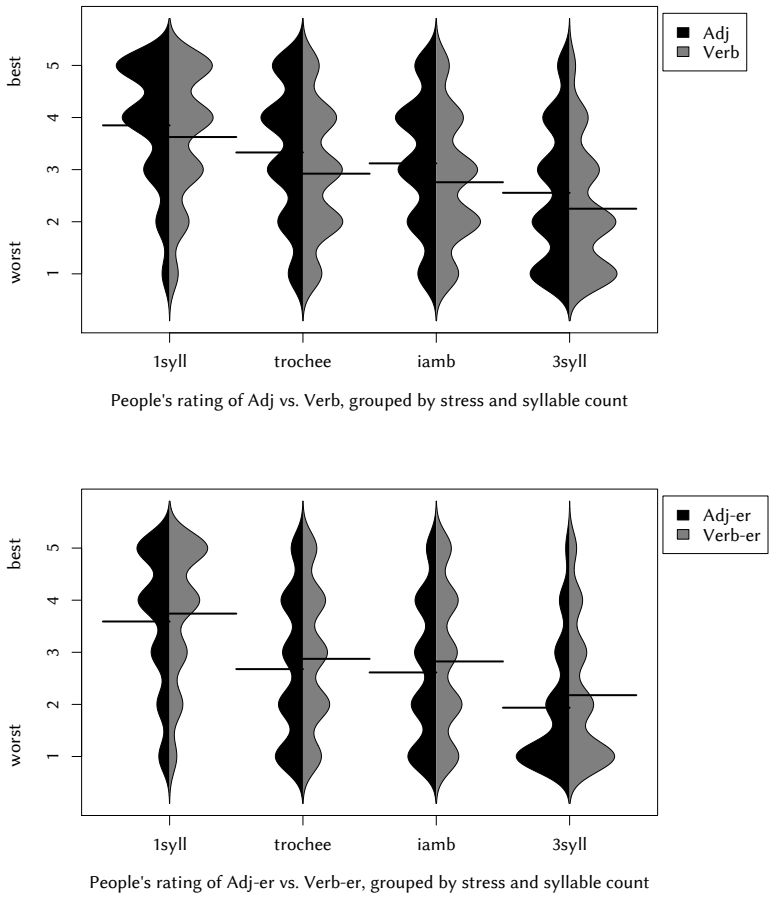


Figure 7: Ratings based on stress location and word length

(t=1.92, p=0.0547). The ratings are higher when an item is presented as an adjective (e.g., "I am fudaddow") than when it is presented as a verb (e.g., "I like to fudaddow"), and the ratings go down with syllable count, so monosyllables such as "zice" receive higher ratings than "ezzep", which receives higher ratings than "fudaddow". Finally, there is a small but significant decrease in ratings for nonce words that receive violations in the phonotactic grammar that was trained on English adjectives that combine with the comparative *-er* suffix in CELEX: thus, the wug [ɪoʊθ] receives 0.00 violations in the adjective *-er* sublexicon grammar, and people rate it on average 4.2 out of 5; compare this to [swæʒ], which gets a violation score of 4.05 from the *-er* sublexicon grammar and 3.4 from our participants. We tested the interaction between this phonotactic violation score and experimental condition (presented as adjective vs. verb), but it was not significant—the phonotactic violation scores in the *-er* grammar affect both verb and adjective ratings the same way.

To analyze the ratings of affixed forms (*fudaddower*, *roather*, *swazher*), we used people's ratings of bases as a predictor, since there was a strong correlation between how

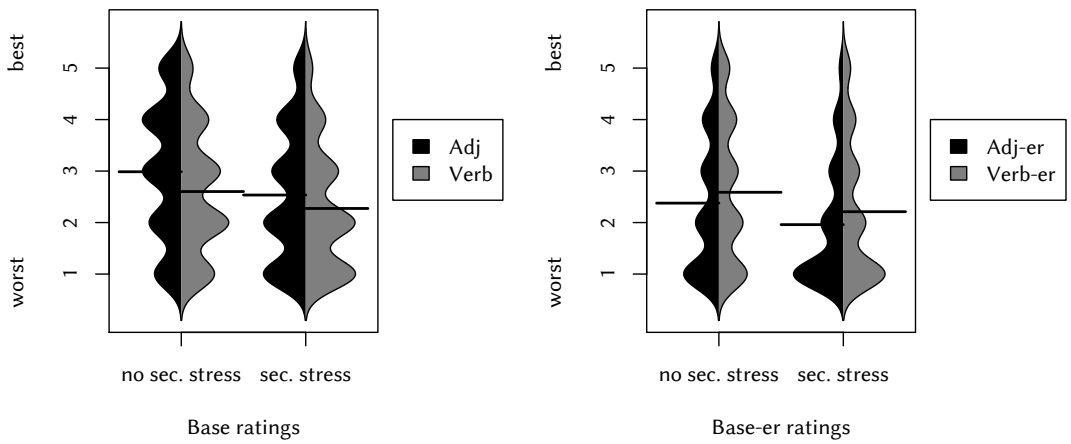


Figure 8: Ratings of nonce words presented as verbs vs. adjectives: presence or absence of secondary stress

	Estimate	SE	t	*p*
Intercept	4.1053784	0.1044277	39.31	¡0.00001
trial number	0.0016690	0.0008686	1.92	0.0547
presented as an adjective	0.3088289	0.0760006	4.06	¡0.00001
syllable count	-0.5544141	0.0518707	-10.69	¡0.00001
phonotactic violations in er-base sublexicon	-0.0227251	0.0037607	-6.04	¡0.00001

max VIF=1.585, max correlation = -.601, AIC=41089, BIC=41264

Table 4: Hierarchical regression model for ratings of bare forms

people rated the base and how they rated the suffixed form. The model included a random by-participant intercept and by-participant slopes for sublexical score, syllable count, and trial number; there were also random by-item and by-frame intercepts and a random by-item slope for experiment name and trial number. Collinearity is reported as VIF, and is low for a fully crossed model. The lower the VIF, the less collinearity between predictors.

As can be seen from the model coefficient summaries for the intercepts, the ratings are overall lower for the affixed forms than for bases. The rating of the affixed form depends largely on the rating of the base (recall Fig. 4): the higher the rating of the base, the higher the rating of the affixed form. The syntactic context of the wug had the opposite effect for bases vs. affixed forms: wugs that were presented as verbs with the *-er* affix (e.g., "an amateur fudaddower") are rated higher than wugs presented as adjectives ("you are even fudaddower"). Violations in the adjectival *-er* sublexical grammar do not have an independent effect on ratings, but when an item has high volations and is presented as an affixed adjective, it is rated lower than when only one of the conditons is met. What this means is that phonotactic violations depress the ratings when the item is presented as

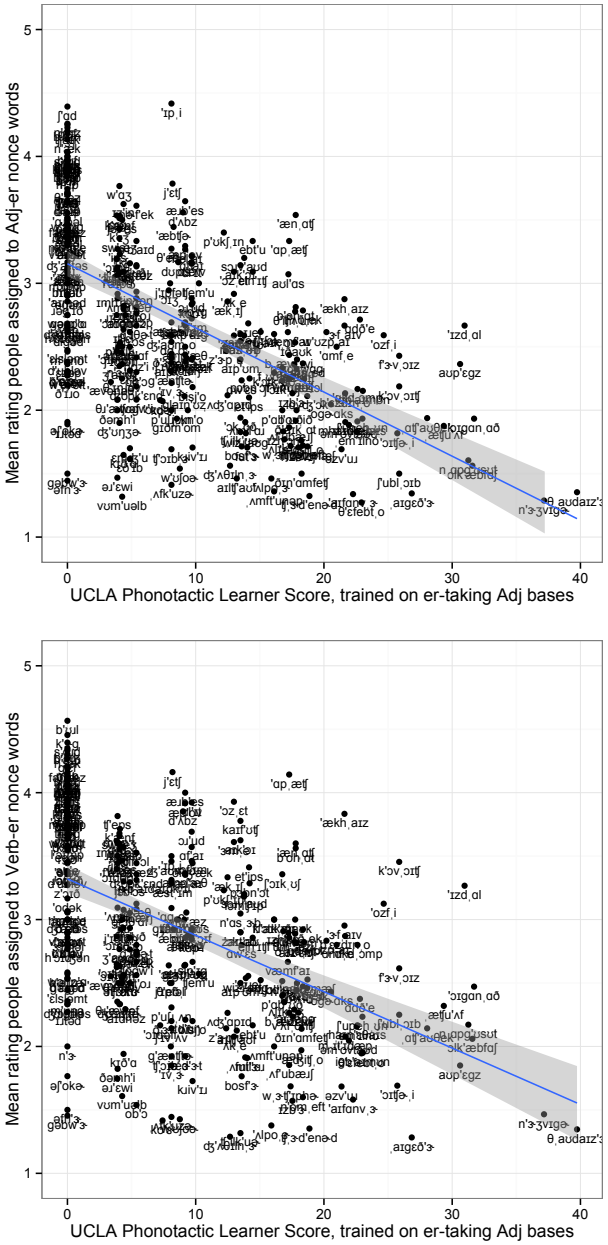


Figure 9: Adj-er and verb-er ratings affected to different extent by violation profile in Adj grammar

	Estimate	SE	*t*	*p*
Intercept	2.3210974	0.0907767	25.57	0.0000
rating of base	0.5138659	0.0077186	66.57	0.0000
presented as adjective-er	-0.3173349	0.0649802	-4.88	0.0001
phonotactic violations in er-base sublexicon	0.0029539	0.0034337	0.86	0.3896
trial number	0.0028168	0.0007079	3.98	0.0000
syllable count	-0.5227117	0.0413162	-12.65	0.0000
er-base-violations : presented as Adj-er	-0.0048867	0.0022204	-2.20	0.0277

max VIF = 2.35; max correlation = -0.548, AIC=39203, BIC = 39393

Table 5: Hierarchical regression model for ratings of affixed forms

an adjective, as expected. Thus, for wugs presented as adjectives, the difference in means between base and base-*er* ratings increases between low-violation and high-violation items, but there is no difference for verbal base and derived forms.

4. Discussion

The results of the experiment are unsurprising given the basic descriptive facts of English morphology: the comparative *-er* suffix is known to be phonologically selective, whereas the nominalizing *-er* suffix does not to show such restrictions; it certainly attaches to trisyllabic and even longer bases (e.g., "nominaliz-er"). What is surprising in the results depends on the theory of phonological selection that one assumes, so let us consider the results from the different perspectives discussed in §2.3.

The traditional phonological analysis of the restriction is that the comparative *-er* attaches to moraic trochees: heavy monosyllables and disyllables with initial stress. This is supposed to be a principled explanation for the effect of syllable count, which most phonologists assume does not arise from actual syllable counting. If the phonological restriction is productive, we would expect to see monosyllables patterning with disyllabic trochees as more acceptable than iambs or trisyllabic words affixed with *-er*. Yet the plot for Adj-er ratings does not show such a grouping—in fact, there is no discernible difference between how people rated the two subclasses of disyllables, iambs and trochees (see the right side of Fig. 7), which were rated lower than monosyllables but higher than trisyllabic words.

When we ran a regression to analyze ratings with "trochee" as a predictor, it was an independent predictor of higher ratings and interacted with syntactic context, but this is hardly surprising—it just makes a cruder distinction than the more detailed predictor of syllable count, which we saw had an effect. Table 6 shows this alternative model of ratings. This model is consistent with the subcategorization frame account of phonological selection: it assumes that trochaicity should have an effect on ratings when the item is suffixed with adjectival "-er".

	Estimate	SE	*t*	*p*
Intercept	1.1620077	0.0647227	17.95	0.0000
rating of base	0.5248372	0.0077084	68.09	0.0000
presented as Adj	-0.4331937	0.0659878	-6.56	0.0000
is a moraic trochee	0.3918052	0.0617322	6.35	0.0000
trial number	0.0027755	0.0007059	3.93	0.0001
presented as Adj : is a trochee	0.1415258	0.0503493	2.81	0.0049

max VIF 1.586, max correlation −0.5340167, AIC=39415, BIC=39567

Table 6: Alternative: subcategorization frame analysis

This model achieves a worse fit to the experimental results (note the higher Akaike Information Criterion number, or AIC), presumably since it misses the finer phonotactic details that affect ratings. In the analysis presented in the previous section, we established that syllable count is correlated with ratings of bases and ratings of derived forms, but sublexical phonotactic violations only have an effect on adjective ratings, not on ratings in general. The "trochee" account is missing this.

The results are also difficult to explain under the view whereby people just use general phonotactic well-formedness to decide on a rating in the experiment. People prefer bare adjectives to verbs, but they prefer suffixed verbs to adjectives. If general phonotactic constraints restrict *-er* affixation in the comparative context, this should be reflected in people's ratings of suffixed forms. But the constraints should affect the ratings of nominalizations, too, and they do not appear to—people rate affixed Verb-er wugs the same as unaffixed Verb wugs.

General phonological constraints do play a role in ratings—thus, people disliked [ɑ]-final nonce words with the *-er* suffix regardless of syntactic context, and differences between Verb-er and Adj-er ratings disappear when the wugs are rhotic-final or end in a non-mid vowel (recall Fig. 6). This result does not adjudicate between approaches to phonological selection, however. Our model allows general phonotactic constraints to affect people's judgments of nonce words, and presumably so would any model that aims to explain people's behavior in experimental settings in terms of grammatical knowledge. But our results do indicate that there are further differences between verbs and adjectives that must be due to affix-specific generalizations, not general English phonology.

5. Conclusion

The phonological selectional restrictions of the English comparative suffix *-er* are well-known, but they are difficult to comprehend given the analytical options available in standard theories of morphophonology. The phonological generalizations cover a range of properties that include the familiar (syllable count and stress position) but also final segment restrictions that hold specially of disyllabic words only. Even when all these generalizations are taken into account, they do not make correct distinctions between English adjectives that do and do not combine with the suffix. Furthermore, English has homophonous

suffixes that either do not impose any phonological restrictions on their bases (the *-er* in *malinger-er*) or impose different, sometimes contradictory ones (*Newfoundland-er*).

We argued that the explanation for these observations is that phonological selection of the English comparative *-er* suffix is really lexical selection. The phonological generalizations emerge from learning phonotactics over the sublexicon of morphemes associated with this specific suffix. This sublexical phonotactic grammar can be used in assessing the likelihood of extending the suffixation rule to novel items. If the new item is phonotactically well-formed in the sublexicon associated with *-er*, it may combine with the suffix. We tested this theory in a nonce word rating study that presented the same wugs as either adjectives or verbs, combining with the two versions of the suffix. People rated the same wugs higher when they were presented as adjectives than as verbs, but when the suffix was attached, the ratings flipped. We argued that this behavior arises because people use affix-specific knowledge rather than general phonological well-formedness.

References

Adriaans, Frans, and René Kager. 2010. Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory and Language* 62:311–331.

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. by Charles B. Chang and Hannah J. Haynie, 1–20. Somerville, MA: Cascadilla Proceedings Project.

Albright, Adam. 2009a. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26:9–41.

Albright, Adam. 2009b. Lexical and morphological conditioning of paradigm gaps. In *Modeling ungrammaticality*, ed. by Curt Rice and Sylvia Blaho, 117–164. London: Equinox Publishing.

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

Baayen, R Harald, Richard Piepenbrock, and H Rijn van. 1993. The {CELEX} lexical data base. {CD-ROM}.

Baker, Mark C. 2003. *Lexical categories: Verbs, nouns and adjectives*, volume 102. Cambridge, MA: Cambridge University Press.

Barr, Dale, Roger Levy, Christoph Scheepers, and Hal Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68:255–278.

Bates, Douglas, and Martin Maechler. 2009. *lme4: Linear mixed-effects models using S4 classes*. URL `http://CRAN.R-project.org/package=lme4`, R package version 0.999375-32.

Becker, Michael. 2009. Phonological trends in the lexicon: The role of constraints. Doctoral dissertation, University of Massachusetts, Amherst, MA.

Becker, Michael, and Blake Allen. 2015. Learning alternations from surface forms with sublexical phonology. Ms. URL `http://ling.auf.net/lingbuzz/002503`.

Becker, Michael, and Maria Gouskova. 2016. Source-oriented generalizations as grammar inference in Russian vowel deletion. *Linguistic Inquiry* 47:391–425.

Belsley, David, Edwin Kuh, and Roy E. Welsch. 2004. *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: Wiley.

Bennett, Ryan. 2016. Output optimization in the Irish plural system. *Journal of Linguistics* .

Benua, Laura. 1997. Transderivational identity: Phonological relations between words. Doctoral dissertation, University of Massachusetts, Amherst, Amherst, MA.

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

Bobaljik, Jonathan David. 2000. The ins and outs of contextual allomorphy. In *University of Maryland working papers in linguistics*, ed. by K. K. Grohmann and Caro Struijke, volume 10, 35–71. College Park, MD: University of Maryland.

Bobaljik, Jonathan David. 2012. *Universals in comparative morphology*. Cambridge, MA: MIT Press.

Booij, Geert. 2011. Morpheme structure constraints. In *The Blackwell Companion to Phonology*, ed. by Marc van Oostendorp, Colin Ewen, Elizabeth Hume, and Keren Rice, volume 4, 2049–2069. Chichester, UK & Malden, MA: Wiley-Blackwell.

Borer, Hagit. 2005. *In Name Only*, volume I of *Structuring Sense*. Oxford: Oxford University Press.

Borowsky, Toni. 1986. Topics in the Lexical Phonology of English. Doctoral dissertation, University of Massachusetts, Amherst.

Boyd, Jeremy Kenyon. 2007. Comparatively speaking: A psycholinguistic study of optionality in grammar. Doctoral dissertation, University of California, San Diego, La Jolla, CA.

Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10:425–455.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Oxford University Press.

Bye, Patrik. 2007. Allomorphy: Selection, not optimization. In *Freedom of Analysis?*, ed. by Sylvia Blaho, Patrik Bye, and Martin Krämer, Studies in Generative Grammar, 63–91. Berlin: Mouton de Gruyter.

Daland, Robert, A. Sims, and Janet Pierrehumbert. 2007. Much ado about nothing: A social network model of Russian paradigmatic gaps. In *Proceedings of the 45th annual meeting of the association for computational linguistics*, 936–943. Prague. URL `http://groups.lis.illinois.edu/amag/langev/paper/daland07socialNetworkRussian.html`.

Di Sciullo, Anna Maria, and Edwin Williams. 1987. *On the definition of word*, volume 14. Cambridge, MA: MIT Press.

Dickson, Paul. 1990. *What do you call a person from . . . ? A dictionary of resident names*. New York & Oxford & Sidney: Facts on File.

Downing, Laura. 2006. *Canonical forms in Prosodic Morphology*. Oxford: Oxford University Press.

Durvasula, Karthik, Ho-Hsin Huang, and Rose Merrill. 2013. Word-internal ambisyllabic consonants are codas. *The Journal of the Acoustical Society of America* 134:4201–4201.

Elenbaas, Nine, and René Kager. 1999. Ternary rhythm and the Lapse constraint. *Phonology* 16:273–330.

Embick, David. 2007. Blocking effects and analytic/synthetic alternations. *Natural Language & Linguistic Theory* 25:1–37.

Embick, David. 2010. *Localism versus globalism in morphology and phonology*. Cambridge, MA: MIT Press.

Embick, David. 2015. *The morpheme: A theoretical introduction*, volume 31 of *Interface Explorations*. Boston/Berlin: de Gruyter Mouton.

Embick, David, and Alec Marantz. 2008. Architecture and blocking. *Linguistic Inquiry* 39:1–53.

Giegerich, Heinz J. 2004. Compound or phrase? English noun-plus-noun constructions and the stress criterion. *English Language and Linguistics* 8:1–24.

Gordon, Samantha. 2014. What do you call a person from .... In *Annual Conference on Phonology*. MIT, Cambridge, MA.

Gouskova, Maria. 2012. Unexceptional segments. *Natural Language and Linguistic Theory* 30:79–133.

Gouskova, Maria, and Michael Becker. 2013. Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language & Linguistic Theory* 31:735–765.

Gouskova, Maria, Sofya Kasyanenko, and Luiza Newlin-Łukowicz. 2015. Selectional restrictions as phonotactics over sublexicons. *Lingua* 167:41–81. URL `http://ling.auf.net/lingbuzz/002673`.

Gouskova, Maria, and Tal Linzen. 2015. Morphological conditioning of phonological regularization. *The Linguistic Review* 32:427–473.

Graziano-King, Janine, and Helen Smith Cairns. 2005. Acquisition of English comparative adjectives. *Journal of Child Language* 32:345–373.

Guion, Susan G., J.J. Clark, Tetsuo Harada, and Ratree P. Wayland. 2003. Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech* 46:403–426.

Halle, Morris. 1973. Prolegomena to a theory of word formation. *Linguistic Inquiry* 4:3–16.

Haspelmath, Martin, and Uri Tadmor. 2009. *Loanwords in the world's languages: A comparative handbook*. Berlin: Walter de Gruyter.

Hayes, Bruce. 2009. *Introductory phonology*. Malden, MA, and Oxford, UK: Wiley-Blackwell.

Hayes, Bruce, and Colin Wilson. 2008. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry* 39:379–440.

Hetzron, Robert. 1975. Where the grammar fails. *Language* 51:859–872.

Hilpert, Martin. 2008. The English comparative—language structure and language use. *English Language and Linguistics* 12:395–417.

Inkelas, Sharon, and C. Orhan Orgun. 1995. Level ordering and economy in the lexical phonology of Turkish. *Language* 71:763–793.

Ito, Junko, and Armin Mester. 1995. The core-periphery structure of the lexicon and constraints on reranking. In *Papers in Optimality Theory II (University of Massachusetts occasional papers in linguistics)*, ed. by Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 181–210. Amherst, MA: GLSA Publications.

Kahn, Daniel. 1976. Syllable-based Generalizations in English Phonology. Doctoral dissertation, MIT, Cambridge, MA. Published by Garland Press, New York, 1980.

Kawamoto, Alan H, and William T Farrar. 1990. Non-obligatory vowel epenthesis in-ed pseudowords: Ambiguity in syllabification resolved by syntax and suffixation. *Language and Speech* 33:137–158.

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

Kiparsky, Paul. 1982. Lexical phonology and morphology. In *Linguistics in the Morning Calm*, ed. by In-Seok Yang, volume 2, 3–91. Seoul: Hanshin.

Kytö, Merja, and Suzanne Romaine. 2000. Adjective comparison and standardisation processes in American and British English from 1620 to the present. In *The Development of Standard English, 1300-1800: Theories, Descriptions, Conflicts*, ed. by Laura Wright, 171–194. Cambridge: Cambridge University Press.

LaFave, Nathan. 2015. The Most Apt Experimental Investigation of English Comparative and Superlative Formation. In *University of Pennsylvania Working Papers in Linguistics*, volume 21, chapter 16. URL `Availableat:http://repository.upenn.edu/pwpl/vol21/iss1/16`.

Leech, Geoffrey N, and Jonathan Culpeper. 1997. The comparison of adjectives in recent British English. In *To explain the present: Studies in the changing English language in honour of Matti Rissanen*, volume 52 of *Mémoires de la Société néophilologique de Helsinki*, 353–373. Helsinki: Memoires de la Societe Neophilologique de Helsinki.

Lewis, Gwyneth, Olla Solomyak, and Alec Marantz. 2011. The neural basis of obligatory decomposition of suffixed words. *Brain and language* 118:118–127.

Lieber, Rochelle. 1980. On the organization of the lexicon. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Lindquist, Hans. 2000. Livelier or more lively? Syntactic and contextual factors influencing the comparison of disyllabic adjectives. In *Corpora galore: Analyses and techniques in describing English*, ed. by John M. Kirk, volume 30, 125–132. Amsterdam and Atlanta, GA: Rodopi.

Lindsay, Mark, and Mark Aronoff. 2013. Natural selection in self-organizing morphological systems. In *Morphology in Toulouse*, ed. by Fabio Montermini, Gilles Boyé, and Jesse Tseng, Selected proceedings of Décembrettes 7, 133–153. Muenchen: LINCOM Europa.

Marantz, Alec. 1988. Clitics, morphological merger, and the mapping to phonological structure. In *Theoretical Morphology*, ed. by Michael Hammond and Michael Noonan, 253–270. San Diego: Academic Press.

Mascaró, Joan. 1996. External allomorphy as emergence of the unmarked. In *Current Trends in Phonology: Models and Methods*, ed. by Jacques Durand and Bernard Laks,

473–483. Salford, Manchester: European Studies Research Institute, University of Salford.

Mascaró, Joan. 2007. External allomorphy and lexical representation. *Linguistic Inquiry* 38:715–735.

Matushansky, Ora. 2002. Tipping the scales: The syntax of scalarity in the complement of seem. *Syntax* 5:219–276.

Matushansky, Ora. 2013. More or better: on the derivation of synthetic comparatives and superlatives in English. In *Distributed Morphology Today: Morphemes for Morris Halle*, ed. by Alec Marantz and Ora Matushansky, 59–78. Cambridge, MA: MIT Press.

McCarthy, John J., and Alan Prince. 1986. Prosodic Morphology. Technical Report #32, Rutgers University Center for Cognitive Science, version of 1996.

McCarthy, John J., and Alan Prince. 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language & Linguistic Theory* 8:209–283.

McCarthy, John J., and Alan Prince. 1993. Generalized Alignment. In *Yearbook of Morphology*, ed. by Geert Booij and Jaap van Marle, 79–153. Dordrecht: Kluwer.

McCarthy, John J., and Alan Prince. 1994. The emergence of the unmarked: Optimality in prosodic morphology. In *Proceedings of the North East Linguistic Society 24*, ed. by Merce Gonzalez, 333–379. Amherst, MA: GLSA Publications.

McIntyre, Andrew. 2014. Constraining argument structure in nominalizations: The case of English-er. *Lingua* 141:121–138.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Aiden. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331:176–182.

Mohanan, K. P. 1982. Lexical Phonology. Doctoral dissertation, MIT, Cambridge, MA. Distributed by IULC Publications.

Mondorf, Britta. 2009. *More support for more-support: The role of processing constraints on the choice between synthetic and analytic comparative forms*, volume 4. Amsterdam: John Benjamins Publishing.

Nevins, Andrew. 2011. Phonologically-conditioned allomorph selection. In *Blackwell companion to phonology*, ed. by Colin Ewen, Elizabeth Hume, Marc van Oostendorp, and Keren Rice, 2357–2382. Wiley-Blackwell.

O'Grady, William, John Archibald, Mark Aronoff, and Janie Rees-Miller. 2005. *Contemporary linguistics*. Bedford St. Martins, 5th edition.

Orgun, C. Orhan, and Ronald Sprouse. 1999. From MParse to control: Deriving ungrammaticality. *Phonology* 16:191–220.

Ott, Dennis. 2011. Diminutive-formation in German. *Journal of Comparative Germanic Linguistics* 14:1–46.

Paster, Mary. 2006. Phonological Conditions on Affixation. Doctoral dissertation, UC Berkeley, Berkeley, CA.

Pater, Joe. 2000. Nonuniformity in English secondary stress: The role of ranked and lexically specific constraints. *Phonology* 17:237–274.

Pertsova, Katya. 2005. How lexical conservatism can lead to paradigm gaps. In *UCLA Working Papers in Linguistics*, ed. by Jeffrey Heinz, Andrew Martin, and Katya

Pertsova, volume 11. Los Angeles, CA: UCLA. http://www.linguistics.ucla.edu/faciliti/wpl/issues/wpl11/papers/Pertsova_Phono6.pdf.

Pesetsky, David. 1979. Russian morphology and lexical theory. Ms., Cambridge, MA. Available at http://web.mit.edu/linguistics/people/faculty/pesetsky/russmorph.pdf.

Pesetsky, David. 1985. Morphology and logical form. *Linguistic Inquiry* 16:193–246.

Poser, William J. 1992. Blocking of phrasal constructions by lexical items. *Lexical matters* 111:130.

R Development Core Team. 2013. R: A language and environment for statistical computing. Vienna, Austria. URL http://www.R-project.org.

Raffelsiefen, Renate. 1999. Phonological constraints on English word formation. In *Yearbook of Morphology 1998*, 225–287. Springer.

Raffelsiefen, Renate. 2004. Absolute ill-formedness and other morphophonological effects. *Phonology* 21:91–142.

Rastle, Kathleen, Matthew H. Davis, and Boris New. 2004. The broth in my brother's brothel: morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin and Review* 11:1090–1098.

Rodríguez, Francisco José Cortés, and María Jesús Pérez Quintero. 2002. On the syntax-semantics interface in word formation: The case of English-er nominalizations. *FUNCTIONAL GRAMMAR SERIES* 213–246.

Rubach, Jerzy, and Geert Booij. 2001. Allomorphy in Optimality Theory: Polish iotation. *Language* 77:26–60.

Safire, William. 1982. My fellow Americanians.

Siegel, Dorothy. 1974. Topics in English Morphology. Doctoral dissertation, MIT, Cambridge, MA.

Smith, Jennifer L. 2011. Category-specific effects. In *The Blackwell Companion to Phonology*, ed. by Marc van Oostendorp, Colin Ewen, Elizabeth Hume, and Keren Rice, volume IV of *Blackwell Reference Online*. Wiley-Blackwell.

Solomyak, Olla, and Alec Marantz. 2010. Evidence for early morphological decomposition in visual word recognition. *Journal of Cognitive Neuroscience* 22:2042–2057.

Steriopolo, Olga. 2008. Form and function of expressive morphology: a case study of Russian. Doctoral dissertation, University of British Columbia, Vancouver, BC.

Turk, Alice. 1994. Articulatory phonetic cues to syllable affiliation: Gestural characteristics of bilabial stops. *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology* 3:107–135.

Vitevitch, Michael S, and Paul A Luce. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40:374–408.

Wiltschko, Martina. 2006. Why should diminutives count? In *Organizing Grammar. Linguistic Studies in Honor of Henk van Riemsdijk*, ed. by Hans Broekhuis, Norbert Corver, Riny Huybregts, Ursula Kleinhenz, and Jan Koster, volume 86 of *Studies in generative grammar*, 669–679. Berlin: Walter de Gruyter.

Wolf, Matthew. 2013. Candidate chains, unfaithful spellout, and outwards-looking phonologically-conditioned allomorphy. *Morphology* 23:145–178.

Wolf, Matthew. 2015. Lexical insertion occurs in the phonological component. In *Understanding allomorphy: Perspectives from optimality theory*, ed. by M. Eulàlia Bonet, Maria-Rosa Lloret, and Joan Mascaró, 361–407. London: Equinox. Available as ROA-912 on the Rutgers Optimality Archive, `http://roa.rutgers.edu`.

Wolf, Matthew, and John J. McCarthy. 2010. Less than zero: Correspondence and the Null Output. In *Modeling ungrammaticality in Optimality Theory*, ed. by Sylvia Blaho and Curt Rice, 17–66. London: Equinox Publishing.

Wurm, Lee H, and Sebastiano A Fisicaro. 2014. What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language* 72:37–48.

Zuraw, Kie. 2000. Patterned exceptions in phonology. Doctoral dissertation, UCLA, Los Angeles, CA.

Zweig, Eytan, and Liina Pylkkänen. 2009. A visual M170 effect of morphological complexity. *Language and Cognitive Processes* 24:412–439.

Maria Gouskova and Suzy Ahn
maria.gouskova@nyu.edu, suzy.ahn@uottawa.ca