ISLANDS OF RELIABILITY FOR REGULAR MORPHOLOGY: EVIDENCE FROM ITALIAN

Adam Albright

University of California, Santa Cruz

The representation of regular morphological processes has been the subject of much controversy, particularly in the debate between single and dual route models of morphology. I present a model of morphological learning that posits rules and seeks to infer their productivity by comparing their reliability in different phonological environments. The result of this procedure is a grammar in which general rules exist alongside more specific, but more reliable, generalizations describing subregularities for the same process. I present results from a nonce-probe (wug) experiment in Italian, in which speakers rated the acceptability of novel infinitives in various conjugation classes. These results indicate that such subregularities are in fact internalized by speakers, even for a regular morphological process.*

1. Overview of the issues. In the course of the debate between connectionist models (Rumelhart & McClelland 1986, MacWhinney & Leinbach 1991) and the dual-mechanism model of morphology (Pinker & Prince 1988, 1994), a substantial body of research has developed describing qualitative differences between regular and irregular inflectional processes.[1] This research program has grown to include an impressive number of domains, including language acquisition, linguistic productivity, lexical recognition tasks, brain imaging, and language pathology (see Clahsen 1999, Pinker 1999, and Ullman 2001 for overviews). Proponents of the dual-mechanism model argue that differences in how regular and irregular forms are produced and processed show that the two must be represented in fundamentally different ways; regular forms are generated by a relatively simple grammar of widely applicable rules, while irregular forms are stored in some type of associative network outside the grammar. Proponents of single-route models such as connectionist models, in contrast, argue that whatever differences there may be between regulars and irregulars emerge as a result of the way that speakers store and generalize over all of the words of their language—regular and irregular—using a single system.

To date, no model—single or dual route—has been implemented that can adequately capture the full range of data. But the large body of experimental evidence about regular/irregular differences poses interesting challenges for all models of morphological productivity. Those who believe in a single mechanism must take on the question of how a unified model can be used to derive seemingly dualistic behavior. Those who believe in a separate rule component, on the other hand, must address the question of how language learners use input data to posit rules, and how rules are evaluated for possible inclusion in the adult grammar.

* Thanks to many people for their helpful criticisms and suggestions, including especially Marco Baroni, Bruce Hayes, Carson Schütze, Donca Steriade, audiences at UCLA, the LSA, and the Workshop on the Lexicon in Phonetics and Phonology at the University of Alberta, and three anonymous referees; all remaining mistakes are, of course, my own. I am also grateful to the participants of the experiment reported here. This material is based on work supported under a National Science Foundation Graduate Fellowship.

[1] The terms regular and irregular have been used in numerous ways; following the literature on English past tense formation, I use regular to refer to productive, default morphological processes, and irregular to refer to unproductive, nondefault processes (even if their formation is straightforwardly describable by rules).

In this article, I explore one approach to morphological rule learning, suggested by Pinker and Prince (1988:130–34) and further developed by Albright and Hayes (2002). This approach compares the morphological behavior of different words in order to hypothesize rules and evaluate their effectiveness in explaining existing words. The goal of learning under this model is to discover the best morphological rules and to provide a quantitative measure of their productivity. For reasons that are explained below, however, the very best rules in this system are not necessarily those that one would choose in a traditional linguistic analysis. In particular, in addition to the most general rules (such as 'suffix *-d* to form the past tense'), the model also finds rules to describe more specific, but more reliable, subregularities for the very same change (such as 'suffix *-d* after fricatives'). I use here the term ISLANDS OF RELIABILITY to refer to such subgeneralizations about phonological environments in which a morphological process is especially robust. This article presents experimental evidence that such subgeneralizations are in fact internalized by speakers and may coexist in grammars alongside the more general rules.

The issue of subgeneralizations for regular and irregular processes has implications for various models of morphology. The dual-mechanism model and more traditional approaches to morphology agree that local subgeneralizations about regular processes are absent from the adult grammar. In other words, although irregular processes may be sensitive to subgeneralizations or neighborhood similarity effects, regular processes should be equally applicable to all new inputs, regardless of their similarity to existing words (Marcus et al. 1995:196). Prasada and Pinker (1993) present results from a nonce-probe (WUG) experiment (Berko 1958) showing that novel regulars are not more acceptable when they are similar to existing regulars; subsequent studies on other languages, such as German, Hebrew, Japanese, and Italian, have found similar results (Marcus et al. 1995, Berent et al. 1999, Fujiwara & Ullman 1999, Say & Clahsen 2000). If speakers are in fact insensitive to the distribution of regulars when applying the regular pattern to novel items, this would be strong evidence that subgeneralizations about regulars are absent from the grammar. But if speakers are sensitive to patterns in the distribution of regulars and can extend these patterns productively to novel items, then we need a way to represent such subgeneralizations in the linguistic system.

Before testing whether existing regulars influence the formation of novel regulars, it is helpful to have explicit quantitative predictions about the nature and degree of their potential influence. Section 2 begins with an outline of a model of morphological learning that makes quantitative predictions about novel formations. This model compares the reliability of different morphological processes in different phonological environments, and uses this information to assign gradient predicted well-formedness scores to novel regulars and irregulars.

2. THE MINIMAL GENERALIZATION MODEL. Much of the work on regular/irregular dissociations has focused on the question of whether the productivity of regulars is sensitive to similarity, or lexical neighborhood effects (Bybee & Moder 1983, Prasada & Pinker 1993, Say 1998, Berent et al. 1999), under the assumption that gradient productivity is best attributed to some sort of analogy that makes direct use of the lexicon. The notion of lexical neighborhoods has proven useful in modeling other on-line tasks, such as lexical access and phoneme disambiguation (Landauer & Streeter 1973, Coltheart et al. 1977, Luce 1986, Newman et al. 1997). However, 'different by *n* features' is a rather crude way to measure phonological similarity, so we should be cautious in drawing conclusions from studies that fail to find lexical neighborhood effects under this definition. Furthermore, even if there is no definition of lexical neighborhoods that adequately describes the productivity of regulars, we would still not have an argument that regulars are handled by a single, context-free rule; we would merely have an argument that more than memory is involved.

The current study employs a model that derives gradient productivity effects using a grammar of stochastic morphological rules, rather than using the lexicon directly. For present purposes, I take it for granted that there is a morphological grammar that is used to project novel forms, focusing instead on how that grammar could be learned, and what rules it might contain for regulars. I assume here that morphological rules specify the changes that are necessary to project one form (e.g. the past tense) from another form (e.g. the present tense): adding an affix, changing a vowel, and so on. In such a system, rules can be expressed as a morphological change (A → B) in a phonological environment (/C__D). For example, the suffixation of the regular past morpheme in English might be expressed as ∅ → *d*/__#.

One method of hypothesizing rules for a morphological change is the MINIMAL GENERALIZATION algorithm, sketched by Pinker and Prince (1988:130–34) and developed further by Albright and Hayes (2002). The premise of this approach is that language learners explore the space of possible phonological environments, looking for those that have especially high reliability for a given change. An environment is said to be an ISLAND OF RELIABILITY when its reliability value is higher than the general reliability of a change. The intuition is that we are looking for statements of the form: 'adding *-áre* works 72% of the time in the general lexicon, but if the verb happens to end in [min], then it works 96% of the time'.

In some languages, the distribution of two competing morphemes is predictable based on the phonological form of the root, such as in the Yidiɲ ergative suffixes: *-du* after consonants and *-ŋgu* after vowels (Dixon 1977). In these cases, collecting statistics about phonological environments can straightforwardly discover the 'true' environments that condition the alternation; the rules adding *-du* after consonants and *-ŋgu* after vowels will have very high reliability, whereas the 'wrong' rules adding *-du* after vowels and *-ŋgu* after consonants would have low reliability. For other languages, like Italian, where class membership is only somewhat less than arbitrary, even the most comprehensive search could not possibly come up with a clean analysis that predicts the correct conjugation class in all cases based solely on phonological environment. But the minimal generalization approach predicts that if there are local subregularities in the data, learners will explore them in an attempt to find the best possible analysis.

The minimal generalization learner uses the lexicon to construct a grammar of (largely redundant) symbolic rules, annotated for phonological context and reliability. It takes as its input pairs of words that stand in a particular morphological relation—for example, ([present], [past]) or ([1sg], [infinitive]), and learns a set of rules to derive one from the other.[2] It does this by carrying out pairwise comparisons of all of the words in the data set and factors them into their change, their shared material, and the unshared residue. The shared material then forms the description of a phonological environment in which the morphological rule is known to apply.

[2] An important issue concerns whether speakers learn to derive forms in multiple directions (e.g. both [1sg] → [infinitive] and [infinitive] → [1sg]), or whether grammars contain rules for only a subset of the possible derivations. I assume for present purposes that the grammar of Italian contains rules assigning verbs to a conjugation class if they have been encountered only in forms that do not reveal this information (such as the 1sg or a loan word from another language). See Albright 2002a for a specific proposal for how learners decide which derivations to include in their grammars.

	RESIDUE	SHARED FEATURES	SHARED SEGMENTS	CHANGE
comparing A:		b	ad	o → are
with B:	di	r	ad	o → are
yields C:	X	$\begin{bmatrix} +\text{cons} \\ +\text{vc} \\ -\text{dors} \\ \text{etc} \ldots \end{bmatrix}$	ad	o → are

FIGURE 1. Minimal generalization to discover environments.

Figure 1 shows the comparison between two Italian ([1sg], [infinitive]) pairs: ([bado], [badare]) 'take care of' and ([dirado], [diradare]) 'thin out'. The two pairs are related by the same change ([o] → [are]), so the remaining parts of the words are compared to locate potentially relevant phonological material in the environment of the change. For simplicity, the description of a phonological environment surrounding a morphological change will be limited to the following components. The first component is the CHANGE LOCATION—in this case, the end of the word. Immediately adjacent to the change location any number of SHARED SEGMENTS, common to both pairs of forms, may be specified. Continuing to move out from the change location, there may next optionally be one partially specified segment, defined by a set of SHARED FEATURES—in this example, a nondorsal voiced consonant. (For the simulations presented in the next section, the capacity for feature decomposition was not employed because it does not materially affect the results for Italian.) To the left of these partially or fully specified segments may also be a RESIDUE (represented by X), allowing any number of unspecified phonemes.

The result of this comparison is a morphological rule, in this case changing the 1sg suffix [o] to the infinitive suffix [are] in a very particular phonological environment. This rule is then compared against the data set in order to see how many words fit the description (the SCOPE of the rule) and how many of those words actually involve the same change (its HITS).

The distributions of scope and hits for different rules can be seen not only as a measure of each rule's popularity but also as a relative measure of its reliability. We can define RELIABILITY as the ratio of a rule's hits to its scope, to provide a 'batting average' for the environment, as in 1.

(1) Definition of reliability

$$\frac{\text{Number of forms included in the rule's structural change } (= \text{ its hits})}{\text{Number of forms included in the rule's structural description } (= \text{ its scope})}$$

For example, the rule environment in Figure 1 covers twenty-one Italian verbs, including [ba'dare], [dira'dare], [gra'dare] 'gradate', [gra'dire] 'be agreeable', [zb^{j}a'dire] 'fade', [per'vadere] 'pervade', and sixteen others. However, only six of these verbs actually belong to the *-are* conjugation. Thus, the rule has a reliability of 6/21 = .286. This ratio is then statistically adjusted, using lower confidence limit statistics (at a confidence level of 75%) following a suggestion of Mikheev (1997).[3] The effect of this

[3] The confidence level is a parameter of the model and can range from 50% to 100%; the value of 75% was chosen in the absence of any independent evidence about the correct value to use. The effect of this parameter is considered in more detail in Albright 2002b, in which it emerges that varying its value does not make a significant difference in the results.

adjustment is to downgrade the reliability scores when the sample size is small—that is, when the rule attempts to cover only a few words. The rationale for this is that we are more certain about things that are widely attested. Reliability calculated in this way is based on type frequency, and token frequency plays no role. This accords with a suggestion in Bybee 1995 that type frequency is probably the most important factor in determining intuitions about novel words. I return to this issue in §4.3.

Albright and Hayes have implemented the minimal generalization algorithm in an automated learner program, which carries out these comparisons iteratively across the entire data set.[4] The result is a list of thousands of ways to describe the phonological environments surrounding the morphological patterns in the language. The list of rules constructed by the minimal generalization algorithm includes many that are extremely specific, covering just a few words. But comparing heterogeneous sets of words also gives rise to descriptions that are extremely general, even to the point of specifying context-free affixation. The context-free rules also have reliability values associated with them, just like any other rule. For instance, in the Italian simulations described below, the 'null description' which allows the affixation of *-áre* to any verb at all (∅ → [are]/ __#) has a hits value of 1,463 (the number of actual *-áre* verbs in the input set) and a scope of 2,022 (the total number of verbs in the input set). This yields an adjusted reliability value of .717. The context-free reliability value of a class may be referred to as its GENERAL RELIABILITY.

When the model is asked to derive a novel form, it tries to apply each rule in the grammar to the novel input. Whenever a rule can apply (when its structural description is met), it is used to derive a candidate output. When deriving the infinitive of a novel Italian 1sg [lavɛsso], for example, some of rules in the grammar would produce the output [lavɛs'sare], while others would produce [la'vɛssere], [lavɛs'sere], and [lavɛs'sire]. Each output is assigned a confidence value, equal to the confidence value of the best rule that derives it. Thus, [lavɛs'sare] receives the confidence value of the best [o] → [are] rule that can apply to [lavɛsso], while [la'vɛssere] receives the confidence value of the best applicable [o] → [ere] rule, and so on. If the novel input falls within an island of reliability for a particular change, then the output employing that change will receive a high score, because it can be derived by a very reliable rule. Novel inputs often fall within more than one island of reliability—for example, adding -*áre* after *ess*, after *ss*, and after voiceless fricatives in general could all be highly reliable rules; it is assumed that the goodness of the output is determined solely by the very best one. If the input does not fall into any island, then the best way a change can apply is by using its general (context-free) reliability. Output scores, like confidence scores, range from 0 to 1.[5]

The fact that the rule with the highest reliability gets to apply means that if the system is forced to choose a unique output from among the possible candidate classes, the class with the highest general reliability is nearly always chosen. It is this mechanism that allows one class to emerge as the globally productive, or regular, class for novel productions. Because there are context-free rules, the regular pattern can apply to novel

[4] The approach as it is described here requires pairwise comparisons of all of the words in the input data, which may seem implausible to some readers because of its inefficiency. It should be noted, however, that the Java implementation of this algorithm runs reasonably quickly—the simulations reported in this paper with input files of 2,022 verbs run in approximately two minutes each on a 450 megahertz PC.

[5] In work on English, we have found that output scores above .6 typically correspond to well-formed outputs, output scores from roughly .2 to .6 correspond to marginal but conceivable outputs, and output scores below .2 correspond to ungrammatical outputs.

words even when they do not resemble existing words to any degree at all. However, if the word happens to contain a phonological environment that is more consistently regular than average, then the regular process can apply with even greater certainty.

Note crucially that high reliability values in this model do not stem from high similarity to any particular words in the lexicon. Rather, they come from finding a large number of morphologically consistent words that can be described by a common structural description, regardless of how general or specific that description is. The statistically corrected reliability values are a quantitative expression of the strength of a morphological process within a phonological environment. These are assumed to correspond to the well-formedness ratings that humans provide for novel forms.

A logical use of the minimal generalization algorithm would be to search for the very best (= highest reliability) rules, which would be included in the adult grammar. But because there are islands of reliability for all changes, including the regular one, in most cases it turns out that the context-free default rule will not be the very best rule in the system. Thus, we have an interesting puzzle: the best rules learned by this model are not the rules that would be chosen by linguists for a traditional linguistic analysis. We therefore must ask what the status of these rules is in the adult language. If speakers are insensitive to islands of reliability for the regular process, as has been claimed in the dual-mechanism literature, then we need to provide the model with a means of diagnosing and discarding them. If, though, we can show that speakers are actually aware of islands of reliability for the regular process, this would be evidence that these rules are retained in some fashion. In order to test this question, I conducted a nonce-probe experiment on conjugation class assignment in Italian, to which I now turn.

3. An experiment to test for islands of reliability.

3.1. The Italian verb conjugations. There are four major verb classes in Italian, illustrated in Table 1. These classes, which I refer to here as the *-áre*, *-ére*, *-ere*, and *-íre* classes, can be distinguished by what vowel they have in the infinitive and whether their stress falls on the root or on the suffix.[6]

VOWEL	STRESS	SUFFIX	SAMPLE ROOT	SAMPLE 1sg. pres	SAMPLE INFINITIVE	GLOSS
[a]	*suffix*	-are	rem-	'remo	re'mare	'row'
[e]	*root*	-ere	frem-	'fremo	'fremere	'quiver'
[e]	*suffix*	-ere	tem-	'temo	te'mere	'fear'
[i]	*suffix*	-ire	dorm-	'dormo	dor'mire	'sleep'

Table 1. The four major Italian verb classes.

The relative distribution of the four verb classes, calculated from the 2,022 verb types found in a spoken corpus of half a million words (de Mauro et al. 1993), is given in Table 2. This table shows that the majority of Italian verbs belong to the default class (*-áre*), but two of the other classes (*-ére* and *-íre*) also contain substantial numbers of verbs.

[6] Traditional Italian grammars refer to only three classes: class 1 is *-áre*, class 2 is *-ere*/*-ére*, and class 3 is *-íre*. I avoid using this classification here because it does not distinguish between *-ere* and *-ére*, which differ significantly in their frequency and phonological contexts—though for a defense of grouping *-ere* and *-ére* as a single class, see Napoli & Vogel 1990. For the purposes of this article, it is possible to ignore the handful of verbs that do not fit neatly into one of the four main classes, such as *porre* 'put' and *produrre* 'produce'.

CLASS	NUMBER	PERCENTAGE
-áre	1463	72.4%
-ere	281	13.9%
-ére	42	2.1%
-íre	197	9.7%

TABLE 2. Relative distribution of the Italian verb classes.

Verbs must always be inflected in Italian (i.e. they cannot appear as bare stems), but not all inflections unambiguously reveal a verb's class.[7] The first singular present indicative ending *-o*, for instance, is the same for all four classes, as can be seen in the fifth column of Table 1. Consequently, if Italian speakers happened to have heard a nonce verb for the first time in the 1st singular, they would not know what class the verb belongs to. But the infinitive suffix is an unambiguous indicator of the verb's class. Therefore, if speakers have heard a verb only in ambiguous inflections, they will need to guess in order to produce an infinitival form. My study simulates this situation experimentally by presenting participants with novel verbs in the first singular and then gathering their intuitions about the various possible infinitives.

Class 1, or the *-áre* class, is the most productive class, and has been identified as the default in the dual-mechanism literature (Say 1998, Say & Clahsen 2000). Various types of data support this claim. First, grammars state that new verbs should belong to the *-áre* class (e.g. Dardano & Trifone 1985). Second, my own experience with native speakers has shown that many (but not all) speakers are also explicitly aware of this fact. Finally, previous studies have shown that the *-áre* class is favored by participants in experimental settings, even for items that are similar to existing *-ere* and *-íre* words (Orsolini & Marslen-Wilson 1997, Say 1998).[8]

As in English, for each of the irregular (nondefault) classes in Italian we can identify at least one robust gang of phonologically related forms that serve as the prototypical members of that class. Davis and Napoli (1994) observe that many *-ére* verbs conform to a prosodic template ($C_{[-son]}VC$), and informal inspection suggests that there may be similar templates for the *-ere* class ($CVC_{[+son]}C_{[+voice]}$, as in ['tendere] 'stretch', ['prendere] 'take', ['mordere] 'bite') and the *-íre* class ($CVC_{[+son]}C_{[-voice]}$, as in [par'tire] 'leave', [men'tire] 'lie', [skol'pire] 'carve'). Davis and Napoli argue that the template for *-ére* verbs has played an active role historically in attracting words into the *-ére* class; similarity studies such as the present one investigate the synchronic force of such templates.[9]

[7] There are four irregular monosyllabic verb stems that end in a vowel and do not receive an additional suffixal vowel in the 3sg or singular imperative: *sta* 'stay', *fa* 'do.3SG/IMPERATIVE', *dà* 'give', and *può* 'be able to'.

[8] It should be noted that the *-íre* class also displays a limited degree of productivity, particularly in the creation of deadjectival verbs, and is treated as regular by Orsolini and Marslen-Wilson (1997). This ability of more than one class to apply productively to novel words is not a problem for the model presented here, in which all inflectional classes are handled by a single mechanism. The issue of the productivity of *-íre* is discussed further in §4.2.

[9] As is perhaps to be expected, none of these templates works perfectly, either as a synchronic description of conjugation class membership in Italian, or as a historical statement about how verbs have changed classes. Leaving aside Davis and Napoli's claims about other Romance languages, even for Italian their template cannot explain every single Italian verb that has remained in or left the *-ére* class (Maiden 1995, Marotta 1997, Wright 1997). The notion that phonological templates may play a role in defining morphological classes is nonetheless valid, even if these particular templates are not the best way to characterize Italian. The model proposed in §2 is meant to provide a more rigorous search and evaluation procedure for locating the synchronically relevant templates.

There are several reasons why Italian is a better test case for similarity effects than English. In English, there are only 180–200 irregular verbs altogether, of which perhaps 160 or so are in current usage. These words are divided into many different classes—Bybee & Slobin 1982 identifies eight major classes and Pinker & Prince 1988 lists around two dozen patterns altogether. In fact, only five or six of the patterns listed by Pinker and Prince contain six or more verbs. With so few irregulars split across so many patterns, a morphological model based on type frequency would predict only weak lexical effects in English, and all novel irregulars should get ratings at floor levels except for a few robust patterns, such as vowel changes from [ɪ] → [ʌ] and [ɪ] → [æ]. Conversely, the regular pattern, which constitutes the vast majority of English verbs, should get ratings at ceiling levels. Therefore, in order to test whether speakers' judgments vary according to the statistics of the lexicon, one must maximize the variability in the statistical predictions by using a language like Italian, which has its irregular verbs consolidated into a smaller number of unproductive classes, each potentially large enough to attract new words. The fact that fewer verbs take the default pattern in Italian means that there is also more room to observe lexical effects in regulars.[10]

Orsolini & Marslen-Wilson 1997 and Orsolini et al. 1998 make a similar argument for using Italian rather than English, and both of these studies show that the default class is not nearly as powerful in Italian as it is in English. However, these previous studies also tested a different regular/irregular distinction in Italian: namely, the formation of past participles, in both productive and unproductive verb classes.[11] Clearly, the relationship between regularity and productivity in languages is complicated by the presence of many morphological classes; the current study attempts to avoid some of this complication by focusing on inflectional class decisions alone instead of past participles, since one class (*-áre*) is clearly productive and must be considered the regular default.

3.2. MATERIALS. The model described above calculates phonological environments based on linguistic structural descriptions, not based on distance to individual existing words. Therefore, the traditional method of constructing novel words for nonce-probe tasks (take existing words and modify them to varying degrees) is not guaranteed to

[10] A number of studies in recent years have discussed cases that seem to act as the default, while containing only a minority of the existing lexicon. The most famous of these is German, in which the *-s* plural is said to be the default despite covering only a small fraction of existing nouns (Marcus et al. 1995). However, there is some contention over the correct analysis of these cases (Clahsen & Rothweiler 1992, Köpcke 1993, Bybee 1995, Clahsen 1999 and replies). The safest strategy, therefore, seems to be to test a case in which the room for lexical effects to surface is smaller, but the status of the default as a default is more secure. If lexical effects can be detected in such a situation despite the smaller room to observe them, then this should provide even stronger evidence of their existence.

[11] In Italian, past participles may be formed by the straightforward addition of a vowel and the suffix [-to], as in /am/ → [am-a-to] 'loved' (*-áre* class), /vend/ → [vend-u-to] 'sold' (*-ere* class), and /dorm/ → [dorm-i-to] 'slept' (*-íre* class), while other past participles are formed less transparently by changing or deleting the final consonant of the root, as in /vintʃ/ → [vinto] 'won' and /perd/ → [pɛrso] 'lost'. If just the past participle suffix is focused on, a novel past participle *boned-u-to* for the root *boned-* could possibly be considered regular because it employs the regular *-to* suffix, while the past participle *bonesso* would be considered irregular. However, in both cases, a novel verb has been assigned to an unproductive conjugation class (*-ere*), so neither form is completely rule-based; and indeed, Say (1998) found that even the 'regular' past participles of unproductive classes behaved as if they were irregular. Furthermore, past participles in Italian are generally only irregular in the unproductive verb classes (particularly *-ere* and *-ére*), so testing regularity with novel past participles is complicated by the relation between participle regularity and conjugation class regularity.

find the words with the highest and lowest predicted ratings under this model. For this reason, nonce words were created in a way that used the predictions of the model more directly.

As mentioned in §3.1, the 1sg present indicative suffix is ambiguous with respect to conjugation class in Italian. A logical way to test whether speakers can use the phonological form of the root in predicting conjugation class is to give them novel words in the 1sg form and ask them to produce or rate an unambiguous form, such as the infinitive.[12] Furthermore, it is necessary to create 1sg forms that fall within islands of reliability for each of the conjugation classes.

The islands of reliability for each conjugation class were found by feeding the model an input file consisting of the 1sg present and the infinitive forms of all 2,022 Italian verb types found in the *Lessico di frequenza* (de Mauro et al. 1993), in phonemic transcription.[13] The model explored the phonological environments for each of the possible morphological changes from 1sg to infinitive: *-o* → *-áre*, *-o* → *-ere*, *-o* → *-ére*, *-o* → *-íre*, and a variety of other minor patterns involving diphthongs (as in [ˈsjɛdo] → [seˈdere] 'sit'), velars (as in [ˈvalgo] → [vaˈlere] 'be worth'), and other isolated irregularities (as in [ˈfatʃ:o] → [ˈfare] 'do', [ˈsɔ] → [saˈpere] 'know', etc.).

A tricky issue arises with the 1sg form of the *-íre* conjugation. In addition to the *-o* suffix, many verbs in this class take an extra [isk] suffix in forms where the root would otherwise be stressed—that is, in the present tense 1sg, 2sg (including the imperative), 3sg, and 3pl. Thus, the 1sg present form of [speˈdire] 'send' is [speˈdisko], not *[ˈspedo]. A limitation of the current implementation of the Albright and Hayes model is that it cannot handle multiple changes simultaneously. As a result, the parser has no way of determining that the [o] in *-isco* is the same [o] as in verbs without the [isk]; the change can be analyzed only as *-isko* → *-íre*, not as a combination of *-isk-* → ∅ and *-o* → *-íre*.

This limitation leaves us with a quandary when calculating phonological neighborhoods of the conjugation classes. On the one hand, the presence of [isk] in the 1sg is a clear giveaway that the stem belongs to the *-íre* class. On the other hand, the focus of the current study is to test whether speakers can predict conjugation class based on the phonological form of the stem, not based on other information that may be available to them. Therefore, I decided to remove the [isk] from all 1sg forms in the *-íre* class. This served as a way of preparsing the data, avoiding an implementational limitation of the current model, and allowing it to compare all of the conjugation classes directly. (I return to this issue in the discussion section.)

The output of the model was a comprehensive list of the structural descriptions surrounding each morphological change—in other words, of the phonological environments included in each conjugation class. This list was sorted by decreasing reliability, in order to identify some islands of reliability for each class. The first useful environments are those that contain many members, belonging exclusively or almost exclu-

[12] It is not clear how often speakers actually find themselves in this situation of having heard a verb only in the 1sg present form. Nevertheless, the 1sg form is relatively common, so it is logically possible that they might be faced with this situation. Furthermore, although the 1sg suffix *-o* is one of the most extreme in neutralizing all four conjugation classes, other, lesser, neutralizations are pervasive in Italian paradigms. In the end, if it can be shown that speakers are able to predict conjugation class based on phonological form, then questions of why they learn to do so, and whether they actually use it in real life, are perhaps secondary.

[13] This input file, along with input and output files for all of the simulations discussed here, the program used for administering the experiment, and the tabulated experimental results are all available from the following web page: http://ling.ucsc.edu/~albright/papers/italian.html.

sively to a single class. For instance, verbs ending in [-ik], [-idz:], [-eʤ:], and [-min] are all well-attested and belong overwhelmingly to the *-áre* class. A longer list of such environments can be found in Appendix B. Similar islands of reliability exist for the other three classes as well; Table 3 shows some islands of reliability for each class.

-áre		**-ere**		**-ére**		**-íre**	
CONTEXT	REL.	CONTEXT	REL.	CONTEXT	REL.	CONTEXT	REL.
X[idz:]	0.985	X[end]	0.937	X[ŋg]	0.191	X[pr]	0.579
X[it]	0.978	X[ud]	0.708	X[oʎʎ]	0.186	X[ap]	0.570
X[eʤ:]	0.959	X[ist]	0.512			X[ib]	0.570
X[min]	0.948	X[im]	0.477			X[un]	0.351
X[ik]	0.847	X[et:]	0.445			X[rt]	0.326

TABLE 3. Some relatively reliable islands of reliability in Italian.

Note: all reliability values are *adjusted* (lower confidence limit, $\alpha = .75$); X represents an unrestricted variable

To test the hypothesis that humans have detailed probabilistic knowledge about phonological environments, I use correlation statistics to test for a relation between the predicted ratings and the actual ratings. It is crucial to include not only environments with high and low reliabilities but also environments with a range of intermediate reliabilities. For this reason, a variety of environments with intermediate predicted values were also selected.

The last step was to create novel verbs belonging to these varied phonological environments. This requires some care, because verbs do not belong to just one single environment under the definition employed here. For example, if a verb with low reliability in *-áre* is needed, one could make up a verb root [tamend-] to belong to the [end] environment. It is possible, however, that even though the [end] environment has a relatively low reliability for *-áre*, the [mend] environment could have a rather high reliability for *-áre*. (This would happen if only ten out of one hundred [end] verbs belong to *-áre*, but of those ten, eight are [mend] verbs, and there are no [mend] verbs in any other class.) In order to avoid this problem, novel forms were designed to ensure that they did not belong to any environments which covered a subset of the forms that the target environment covered. In other words, if the target environment was [end], no novel verbs were created in [mend] because there is a pre-existing [mend] environment with potentially different reliabilities.

Forty novel verb roots were created in this way; a complete list is given in Appendix A. Four phonotactically unusual or ill-formed verb roots were used from Say's 1998 study: *zuisd-* [dzuizd], *pebcr-* [pebkr], *ghirilb-* [girilb], and *grobm-* [grobm], yielding a total of forty-four novel verbs. As a control, ten existing verbs were also included, representing each of the four classes, and spaced evenly throughout the test items. The existing control verbs were not included in the statistical analyses presented below. The test items were pseudo-randomized so that real verbs occurred approximately every seven to eight items, and completely impossible verbs occurred every ten verbs or so. Also, similar novel verbs were spaced out throughout the list, so they did not occur in close proximity to one another.

3.3. METHOD. The list of fifty-four verbs (forty-four novel, ten real) was presented to twenty-seven adult native speakers of Italian in an acceptability ratings task. All of the participants were from Central or Northern Italy and had completed at least some college education in Italy. Participants were paid for their participation.

The ratings were collected using a computer-administered survey. Stimuli were presented auditorily over headphones, and appeared simultaneously in orthographic form on the computer screen. Test items were first presented in the ambiguous 1sg present form, in a frame sentence: *Oggi* [verb-1sg.pres] *con mio fratello* 'Today I am [verb]-ing with my brother'. Participants then rated the phonological goodness of the root using a slider bar to the right of the sentence. For this part of the task, participants were instructed to judge how typical or 'Italian-sounding' the verb was. The slider bars were marked with a seven-point scale (1 = worst, 7 = best), but the slider itself was continuous, and participants were instructed that they should feel free to slide the handle to any position along the bar.

The 1st singular rating provided a baseline rating for each verb root and also gave participants a chance to internalize the verb root independently of its verb class. Participants then heard the same verb in the infinitival form for each of the four verb classes, in the frame sentence (*Mi piace* [verb-inf] 'I like to [verb]'). For each version of the infinitive, participants were told to rate how good the form would be as the infinitive of the novel verb. The sentences with infinitives were revealed progressively and remained on the screen until all four infinitives had been rated, as seen in Figure 2.

FIGURE 2. Ratings screen for the root and all four classes.

Each participant saw the four possible infinitives in the same order for all items; they were told that the order would be constant but arbitrary. Half of the participants received the infinitives to rate in the order *-áre*, *-ere*, *-ére*, *-íre* (which also corresponds to the canonical order of the verb classes in grammars), and the other half received them in the opposite order. (These orders were chosen because they correspond to the conventional order found in Italian grammars and can be used without suggesting any particular agenda.) While participants were progressing through the series of four infinitives for a given root, they were free to go back and revise their other ratings for that root. They were discouraged from doing too much readjusting, however, and were instructed to provide their first impression of each form. When all four infinitives had been presented and rated, a NEXT button appeared, allowing participants to move on to the next form. It was not possible to go back and readjust ratings for previous items once the NEXT button had been clicked. Due to limitations in the software, all participants rated the test items in the same order; however, because the test items were nonsense words, any possible priming effects due to lexical access should have been less than if the test items had been real words.

In order to make sure that participants were paying attention to the verbs, they would occasionally receive a multiple-choice question when they clicked the NEXT button. They would be presented with the question *Cosa faccio oggi?* ('What am I doing today?'), and they would have to identify the most recent nonce verb from a list of four phonologically similar nonce verbs. There were twelve such quiz questions; it was decided ahead of time that any participant who missed two or more quiz questions would be excluded from the analysis, since this would imply that they were not paying attention for up to a fifth of the session. Four participants were excluded for this reason, each missing exactly two quiz questions.

3.4. RESULTS.

VARIABILITY OF RATINGS ACROSS CLASSES: Under the strongest possible interpretation of the dual-mechanism model, one might expect a complete dissociation between ratings of the *-áre* class and ratings of the other three classes: the rule that produces the *-áre* class is equally applicable to all novel items and should thus be consistently acceptable, regardless of the phonological form of the root. The other classes, in contrast, would vary in acceptability depending on how similar the novel root is to existing irregular roots.

One way to measure the consistency of acceptability for a verb class is by looking at the standard deviation of the ratings for novel verbs in that class. Table 4 shows that the size of a standard deviation of mean ratings for the forty-four novel verbs is comparable across all of the classes; the only exception is the third class, *-ére*, which exhibits a floor effect because the mean rating for this class is so low. Thus, the strongest version of the dual-mechanism model is not supported; judgments about the default class (*-áre*) vary with the phonological form of the verb, to about the same extent that judgments about the nonproductive classes do. The *-áre* forms that received high ratings included [vegeʤ:'are] (mean rating 6.0), [g^warmi'nare] (5.6), and [frak^{w}i'nare] (5.3). Some *-áre* forms that received relatively lower ratings included [adduŋ'gare] (4.2), [funa'dare] (4.2), [spina'dare] (4.4), and [ʃiŋka'dare] (3.8). Among the lowest-rated *-áre* forms were also the four phonotactically ill-formed words, [dzuizdare] (4.2), [grobmare] (3.6), [blinsare] (4.7), and [braipsare] (4.6).

CLASS	MEAN	STD. DEV.
-áre	4.9	2.7
-ere	3.4	2.9
-ére	2.4	2.4
-íre	3.8	2.8

TABLE 4. Mean and standard deviation of mean ratings for novel verbs in the four classes.

This finding alone is not surprising, however. In fact, many previous studies, including Prasada & Pinker 1993, Marcus et al. 1995, and Say 1998, have found variability in the acceptability of novel regular forms. Such variability may be incompatible with the strongest interpretation of the dual-mechanism model, but it is nonetheless compatible with a weaker interpretation, under which the default rule may potentially apply to all input forms, but regular outputs may be partially blocked or degraded by other factors, such as phonological well-formedness and competition from strong irregular outputs. Crucially, this weaker interpretation of the dual mechanism model still precludes the possibility that the acceptability of novel regulars could be improved by attraction to existing regulars. In the next three sections, I consider the following questions: (1) Can the observed differences between novel *-áre* infinitives be explained by the influence of existing *-áre* verbs? (2) Can the influence of existing *-áre* verbs be teased apart from the influence of phonological well-formedness? and (3) Can these two factors be teased apart from the effect of competition from irregular patterns?

CORRELATION TO PREDICTED VALUES: So far, it has been shown that not all novel *-áre* forms were rated as equally acceptable. But is this due to the influence of existing *-áre* verbs in particular phonological environments or is it random experimental variability? To test the idea that speakers are aware of islands of reliability not only for irregulars but also for regulars, one can look for a correlation between the acceptability ratings from participants and the environment reliability values whose calculation was described above.

Table 5 shows that for all classes except *-ére*, there is a significant correlation between acceptability and environment reliability. The *r* values listed in Table 5 are not extremely high, but this may be due to the rather crude definition of reliability employed here. The important thing to notice about Table 5 is that the correlation to environment reliabilities for *-áre* is nonzero and highly significant. In addition, although the correlation exhibited by *-áre* is not the strongest of all four classes, it is also not the weakest. This is evidence that environment effects are not limited only to unproductive classes; the regular, productive class is also sensitive to them.

CLASS	CORRELATION	SIGNIFICANCE
-áre	$r(42) = .61$	$p < .0001$
-ere	$r(42) = .75$	$p < .0001$
-ére	$r(42) = .26$	n.s.
-íre	$r(42) = .31$	$p < .05$

TABLE 5. Correlation between acceptability ratings and environment reliabilities.

Note: all *p*-values for correlations are two-tailed

This finding is consistent in some ways with several previous studies. Orsolini and Marslen-Wilson (1997) found that regular and irregular patterns are equally susceptible to analogy with similar verbs in priming and elicitation tasks, and Orsolini, Fanari, and Bowles (1998) present similar evidence from children. Say (1998) claims that the *-áre* class showed no similarity effects in her study; however, she did not include test items that were specifically designed to be similar or dissimilar to existing *-áre* verbs, so her study did not really test this directly. Say did find that her participants were more likely to volunteer novel forms in the *-ere* and *-íre* classes when the roots were very similar to existing verbs in those classes, which is consistent with the results in Table 5. Note that the previous findings are not identical to those reported here, because earlier studies have all used similarity and lexical neighborhoods, instead of reliability within a phonological environment, to predict lexical effects. The results are similar, however, insofar as they reflect an influence of existing lexical items on novel words.

FACTORING OUT PHONOLOGICAL WELL-FORMEDNESS: The results of the previous section are compatible with the hypothesis that speakers are aware of islands of reliability not only for unproductive, irregular morphological classes, but also for the regular, default class (*-áre*). Nothing presented so far precludes an alternative interpretation, however, one more compatible with the dual-mechanism model; namely, that the differences between novel *-áre* forms are due to phonological ill-formedness and competition from irregular outputs. Indeed, although the novel verbs in this study were for the most part designed to be as well-formed as possible Italian verbs, some of the lowest-rated *-áre* forms were the four ill-formed verbs ([dzuizdare], [grobmare], [blinsare], [braipsare]). Is it possible that all of the differences between novel *-áre* forms can be attributed to small differences in the phonological well-formedness of the roots, even for those verbs that were intended to be well-formed?

To rule out this confound, some way of factoring out the influence of phonological well-formedness on the morphological judgments is needed. Recall that in addition to rating novel infinitives, participants also provided phonological well-formedness ratings of novel verbs in the 1sg form. These can be used to perform a partial correlation, factoring out one variable (phonological well-formedness) and then testing for a correlation between the remaining variables (morphological acceptability and the neighborhood reliabilities). The results of this partial correlation are given in Table 6. By comparing Table 6 with the simple correlations in Table 5, it can be seen that even though a very small amount of the variation goes away when phonological well-formedness is factored out, the bulk of the effect cannot be dismissed in this way. This result can be contrasted with that of Prasada and Pinker (1993), who found that once phonological well-formedness was factored out, they were left with no unambiguous effect of existing regulars on the acceptability of novel regulars. In the present case, only a small proportion of the differences between novel *-áre* forms came from the phonological strangeness of verbs like [grobmare].

CLASS	CORRELATION	SIGNIFICANCE
-áre	$r(42) = .57$	$p < 0.0001$
-ere	$r(42) = .76$	$p < 0.0001$
-ére	$r(42) = .26$	n.s.
-íre	$r(42) = .31$	$p < 0.05$
(overall)	$r(174) = .78$	$p < 0.0001$

TABLE 6. Partial correlation between acceptability and neighborhood reliability (controlling for phonological well-formedness).

In order to provide a more stringent test of the relative effects of morphological and phonological well-formedness, a linear regression (stepwise, mixed) was performed, trying to predict the acceptability ratings of novel *-áre* forms. The two factors included were the phonological well-formedness of the root and the model's environment reliability values for *-áre* forms. If ratings for novel *-áre* forms were based mainly on phonological well-formedness, then the root ratings would be entered first in the model and would explain more of the variance. In fact, the morphological predictions based on environment reliabilities were entered first ($F(1) = 19.64$, $p < 0.0001$) and accounted for 38% of the variance. The phonological well-formedness factor was entered second ($F(1) = 16.99$, $p < 0.001$) and accounted for an additional 18% of the variance, yielding a total model r^2 of .56. Thus, it appears that both morphological and phonological considerations played highly significant, but independent, roles in influencing ratings of novel *-áre* forms. I discuss this finding for Italian in relation to Prasada and Pinker's claims for English in §4.4 below.

COMPETITION FROM IRREGULARS: THE TRADE-OFF HYPOTHESIS: We have seen thus far that ratings of the default *-áre* class are in fact susceptible to lexical effects and do not merely reflect the phonological strangeness of the novel words. There may, however, be another way to interpret these results, compatible with the dual-mechanism hypothesis. Perhaps people are not sensitive to phonological environments for *-áre* forms, but they give certain forms lower ratings when the competing classes receive higher ratings. This could be called the TRADE-OFF HYPOTHESIS, since it assumes that people do not rate the different conjugation classes independently.

To test this, first observe that ratings of the *-áre* class do show an inverse correlation to the sum of the ratings from the other three classes ($r(42) = -.30$, $p < 0.05$), but it is not an especially strong one. In order to provide the strongest test of the hypothesis that environment reliability effects of the *-áre* class are being observed, and not the combination of phonotactic well-formedness and trade-offs with the other classes, the multiple regression from above was repeated, but with the observed ratings of all three other classes (*-ere*, *-ére*, and *-íre*) included as possible predictor variables. This time, a stepwise forward model was chosen, in order to force all other factors to be considered before the predicted *-áre* environment effects. As before, the morphological predictions based on environment reliabilities were entered first ($F(1) = 8.52, p < 0.01$), followed by phonotactic well-formedness ($F(1) = 23.22, p < 0.0001$). The observed ratings of two of the other three classes were entered last, both at nonsignificant levels (*-ere*: $F(1) = 3.803, p = 0.06$; *-íre*: $F(1) = 2.853, p = 0.10$). Trade-off effects with these two classes accounted for an additional 8% of the variance in *-áre* ratings, yielding a total model r^2 of .64. Thus, although a small amount of trade-off behavior can be observed in the ratings of the *-áre* class, it does not appear to be the major determinant in raising or lowering the ratings of novel *-áre* forms.

The predictions of this complete model for *-áre* ratings, taking neighborhood effects, phonotactic well-formedness, and trade-off behavior into account, are plotted against the observed ratings of *-áre* forms in Figure 3.

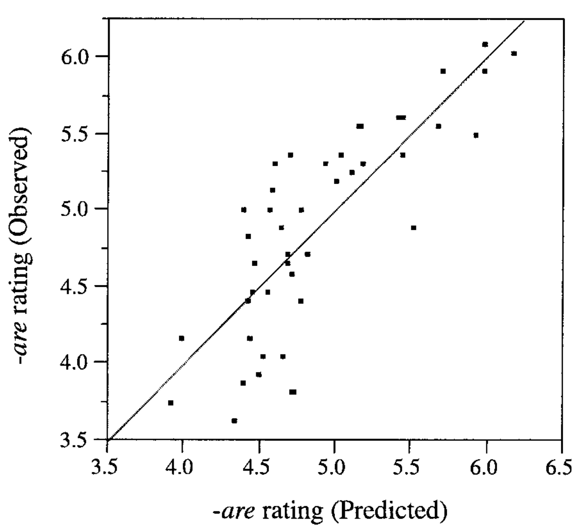

FIGURE 3. Complete model predicting acceptability of *-áre* forms, using morphological predictions, phonotactic well-formedness, and competition from other classes.

4. DISCUSSION. The results presented here contradict a previous set of findings in support of the dual-mechanism hypothesis—namely, that regular inflectional processes are applied in a way that is insensitive to the influence of existing regular forms. A number of questions remain, however.

4.1. DIFFERENCES BETWEEN THE CONJUGATION CLASSES. The correlations in Table 5 show a surprising result: the minimal generalization model does a much better job at capturing differences for the default conjugation class (*-áre*) and one irregular class (*-ere*) than it does for the other two irregular classes (*-ére* and *-íre*). Yet all models of morphological productivity agree that the irregular classes are highly sensitive to the phonological shape of the novel root. Why would the model fail on these particular irregular classes?

For the *-íre* class, I believe that the answer has to do with the *-isc-* affix, mentioned in §3.2. To allow the model to explore the class membership of verb roots based on their phonological shape alone, I removed this suffix from the input data. This was done to overcome a parsing limitation in the implementation of the model, thereby allowing all conjugation classes to be analyzed in parallel fashion. In real life, however, speakers clearly do have more information available to them than just the phonological form of the root. Of the 197 *-íre* verbs in the database, 131 of them (or 67%) take the *-isc-* suffix. Furthermore, verbs that do not take the *-isc-* suffix frequently exhibit other irregularities in the 1sg form, as in [vɛŋgo] ~ [venire] 'come', [diko] ~ [dire] 'say', [ɔdo] ~ [udire] 'hear', [ɛsko] ~ [uʃ:ire] 'exit', etc. What is surprising, therefore, is how high the ratings were for novel *-íre* forms, considering that none of them contained a characteristic marker.

Similar considerations hold for the *-ére* class. In this class, there is no equivalent of the *-isc-* suffix for stressed forms, but the stem alternations are more diverse—[tɛŋgo] ~ [te'nere] 'hold', [ri'maŋgo] ~ [rima'nere] 'remain', [pɔsso] ~ [po'tere] 'be able to', [sjɛdo] ~ [se'dere] 'sit'. In fact, 43 of the 58 *-ére* verbs in the database exhibit such alternations. The nonce verbs in this study did not contain such alternations, since the goal was to test for lexical effects on the default (*-áre*) class, not the irregular classes. The ratings for novel *-ére* forms were, on average, lower than the ratings for novel *-íre* forms. Nevertheless, it is noteworthy that they were rated as plausible at all, given that they did not exhibit any alternations.

Could it be that speakers were ignoring the lack of alternations when making their ratings? In particular, instead of answering the question, How good would Y be as the infinitive of X?, might they have been answering the simpler question, How good would Y be as an infinitive of Italian? This is an especially plausible hypothesis for the half of the participants that rated the *-ére* and *-íre* classes last, since the 1sg form was presented only once at the beginning.

In an attempt to answer this, I created a new input file of Italian verbs, in which the 1sg form was replaced with the uninflected verb stem, as it appears in the infinitive. Each input pair therefore consisted of a bare verb stem and its infinitive form: [ven] → [venire], instead of [veŋgo] → [venire]. This is equivalent to asking the model, How likely is [ven] to be a member of the *-íre* class?, regardless of whether [ven] shows up as [ven], [veŋg], [vjen], or something else in the first singular. This was intended to simulate a participant who was ignoring the 1sg forms and only paying attention to the infinitives. The results are shown in Table 7.

CLASS	CORRELATION	SIGNIFICANCE
-áre	$r(42) = .31$	$p < 0.05$
-ere	$r(42) = .79$	$p < 0.0001$
-ére	$r(42) = .44$	$p < 0.01$
-íre	$r(42) = .39$	$p < 0.01$

TABLE 7. Model based on stem as found in the infinitive.

As can be seen from the table, the model is somewhat more successful for the *-ére* and *-íre* classes when 1sg alternations are ignored. The correlation for the *-ere* class remains consistently high, while the correlation for *-áre* is lower but still significant at the 0.05 level. These results are essentially the same for ratings from both presentation orders (*-áre* first vs. *-íre* first). If participants had been considering the 1sg form at the beginning of the ratings, when they had recently heard the novel verb in the 1sg form, but not at the end, after they had heard three other intervening infinitives, then one might expect that this stem-only simulation would be better for classes when they came last, and worse when they came first.

Another feature of *-ére* and many *-íre* verbs that may be relevant is their high token frequency. A modified version of the minimal generalization model that takes token frequency into account is discussed below in §4.3; it turns out, however, that allowing token frequency to influence this particular model does not improve its predictions for these two classes.

Thus, the remaining puzzle is that participants' ratings for the *-ére* and *-íre* forms can be modeled to a certain extent, but not as well as the other classes. In the end, this may simply be a result of the fact that other factors (such as *-isc-* or stem alternations) already serve as much better cues to membership in these classes. The task of evaluating novel *-ére* and *-íre* forms without these cues was probably a rather strange one, and participants may well have adopted some strategy for rationalizing these forms. It is probably not possible to determine what this strategy was exactly, but for the present purposes, it suffices that Table 7 provides at least one plausible interpretation of these ratings. In other words, it does not appear that the ratings were completely bizarre or unexplainable for any of the conjugation classes.

4.2. The productivity of -ire. Another mystery in the analysis of Italian is why the *-íre* class should be marginally productive, especially in the realm of deadjectival verbs. This productivity is especially perplexing considering that the *-ere* class actually has a somewhat higher type frequency and would therefore be expected to be more productive than *-íre*. I believe that the minimal generality approach can shed some light on this problem.

An issue that has generally not been addressed by computational models of morphological productivity is how the productivity of a suffix may depend on the morphosyntactic or semantic domain. To give one other example, Marcus et al. 1995 found that the *-s* plural was more strongly favored for some implied meanings of novel nouns than for others. The data fed to the minimal generalization learner in this study involved only the phonological form of the word. In principle, however, the same procedure of comparing shared features and computing reliability values in the presence of different conditioning environments could be used to learn that a particular process applies quite regularly within a specific syntactic or semantic subdomain. For example, if it could be stipulated that there is a feature marking verbs as being related to adjectives, then it might be discovered that *-íre* is much more reliable in the $/X]_{Adj}]_{Stem}$— environment than in more general $/X]_{Stem}$—. Alternatively, the change itself could be described as a combination of syntactic and phonological changes: *[adj.] → [verb] · [-íre]*. This strategy could also be used to discover that a process is restricted to specific gender contexts, transitivity contexts, and so on.

4.3. The basis of reliability. The definition of reliability used here is based on type frequency, not token frequency. This accords with Bybee's suggestion (1995) that type frequency has the greatest influence on people's intuitions about novel words. The correlation results that I have presented here support the idea that a measure based on type frequency does match human judgments fairly well. Furthermore, the definition of reliability given above, along with the statistical adjustment using lower confidence limits, are intuitively rational ways that a language learner might make use of type frequencies. But this is only one way among many to count patterns in the lexicon. Might even better results be obtained by counting in some other way?

To test various ways of using type and token frequency, I modified the Albright and Hayes learner program to use a number of different measures as the basis of predicted values. In particular, I tried several variants of the reliability function, using different confidence limits in the statistical adjustment (75%, 90%), and also using no statistical adjustment at all. I also tried a definition of environments that included the featural term, and a version of reliability that was weighted for the length of the shared string term in the structural description (reliability $\times 1.2^n$ for n shared segments), under the hypothesis that significant overlap is more likely to induce analogy to existing forms. In addition, I included a version of reliability based on the tokens of forms covered by the neighborhood description (i.e. its TOKEN HITS) and the tokens of forms to which it applies (its TOKEN SCOPE). Finally, I included a measure that combined relative type and relative token reliability by multiplying the two together.

These comparisons, given in Table 8, show that all measures based on frequency (type or token) perform relatively well. Although different measures perform differently on different classes, there does not appear to be much evidence favoring one single measure as the most accurate model of human intuition. At a broad level, measures that incorporate type frequency do outperform those based on token frequency (the last two lines in Table 8), but only for some classes.

RELIABILITY MODEL	**-are**	**-ere**	**-ère**	**-ire**
adjusted reliability, $\alpha = .75$	0.6137	0.7544	0.2504	0.3126
adjusted reliability, $\alpha = .90$	0.6256	0.7475	0.2522	0.1956
raw reliability	0.5834	0.7617	0.2558	0.3562
with features	0.4928	0.7386	0.3763	0.3310
weighted by length	0.6141	0.7412	0.1701	0.3095
token frequency	0.3423	0.7698	0.3183	0.1387
weighted by token frequency	0.4429	0.7378	0.2941	0.1813

TABLE 8. Effect of reliability basis on correlation to human ratings.

Two reasons come to mind why type and token frequency perform so similarly here. First, the test items were not designed explicitly to tease apart the influence of type vs. token frequency, so the predictions of the two are quite similar for this batch of test items ($r(174) = .79$, $p < .0001$). To address this question directly, I would need to include test items modeled on existing words with exceptional patterns but high token frequency. (For instance, the extremely common verb *faccio* [ˈfatʃ:o] 'do.1SG' has an irregular infinitive *fare* [ˈfare].) Second, the corpus for the token frequencies (de Mauro et al. 1993) was quite small (500,000 words), so low-frequency words were not adequately differentiated to make subtle predictions based on token frequency.

A larger question that cannot be answered here is whether another model of lexical effects, such as a connectionist model or the ANALOGICAL MODEL OF LANGUAGE (Skousen 1989, Eddington 2002), could also model the Italian data. An important difference between such models and the Albright and Hayes model is that the former typically do not assume that grammatical knowledge has been precompiled ahead of time, while the latter sees the exploration of environments as part of the grammar acquisition process. An off-line ratings task cannot help us to decide whether linguistic generalizations are precompiled or done on the spot in response to the novel words. This is an important issue, though, that should be investigated further by other means.

4.4. THE DUAL MECHANISM MODEL. The results of this study contradict a growing body of literature arguing that generalization of regular morphological processes is insensitive to the influence of existing regulars. This claim has taken several forms.

Pinker and Prince (1988) point out that at least for English past tenses, the regular pattern applies to a broad assortment of verbs, so there is no such thing as a prototypical regular. They go on to argue that the distributional difference between regulars and irregulars is mirrored by a behavioral difference in their productivity—in particular, although irregular patterns are very sensitive to the phonological form of the root, regular inflections are claimed to be free from similarity or neighborhood effects (1988: 117).

In subsequent connectionist work on English past tenses, it has been argued that one way in which networks can generalize the regular pattern is precisely by exploiting the distributional difference between scattered regulars and clustered irregulars (Hare & Elman 1995, Nakisa et al. 1997). In response to this, the distributional claim about regulars has been refined in the dual-mechanism literature. Ullman (1999) observed that although English regular verbs do not cluster in any single part of phonological space, some regions of phonological space are in fact more consistently regular than others, when one uses a metric comparing the ratio of regular neighbors (friends) to irregular neighbors (enemies). Similarly, Berent et al. 1999 claims that in Hebrew the regular (masculine) plural falls into several well-defined phonological regions. In both cases, experimental evidence is offered to show that the regular pattern is nonetheless immune to neighborhood effects. The revised claim, then, seems to be as follows: languages may differ in how uniformly regulars are scattered across phonological space, but even when regulars fall into coherent clusters, speakers are insensitive to this distributional fact when applying the default to novel words.

In the present study, I have shown that in Italian, as with English and Hebrew, regular words are not distributed evenly throughout the phonological space; there exist phonological islands of reliability in which the regular class is especially strong. Furthermore, it appears that speakers are sensitive to these islands when generalizing the *-áre* class to novel verbs, in a way that cannot be reduced to the independent effects of phonological well-formedness and competition from irregulars. Why do existing regulars play a significant role in the generalization of the regular pattern in Italian, but not in English (Prasada & Pinker 1993), German (Marcus et al. 1995), or Hebrew (Berent et al. 1999)?

It is possible that previous studies have failed to find coherent lexical effects in the ratings of novel regulars because of the model of neighborhood effects that they have assumed. In these studies, novel items were classified by their similarity to existing words, measured by the number of phoneme changes needed to change the closest existing word into the novel word. The current study, in contrast, categorizes novel items based on the phonological environments that they fall into. A natural question is whether the definition of environment reliability can do a better job modeling Prasada and Pinker's data.

To test this, I submitted to the minimal generalization learner a database of 2,181 ([present], [past]) pairs of English verbs, in phonemic transcription. This database was based on the file used for the English past tense simulation by MacWhinney and Leinbach (1991),[14] augmented slightly to include all of the irregular verbs of English. The list of environment reliabilities was then used to derive predicted ratings for all of Prasada and Pinker's pseudo-regular and pseudo-irregular nonce words. As above, correlations were then computed separately for regulars and irregulars, with phonological well-formedness (root) ratings partialed out. The result was a significant correlation

[14] This file can be downloaded from http://psyling.psy.cmu.edu/papers/.

not only for ratings of irregular forms ($r(58) = .57$, $p < 0.0001$) but also for ratings of the regular forms ($r(58) = .39$, $p < .01$). In a linear regression model, root rating was entered first ($F(1) = 26.535$, $p < .0001$), but the model's predictions based on environment reliability were also entered significantly ($F(1) = 9.529$, $p < .01$).

It is beyond the scope of this article to provide a comprehensive analysis of English, but this discussion should suffice to show that even in Prasada and Pinker's data, there appear to be island-of-reliability effects for regulars. Thus, I do not see any reason to believe that English and Italian are qualitatively different in this respect.[15]

These results pose a problem for the dual-mechanism approach, since it is not clear how existing regulars could influence novel regulars in this model. The effect could not come from the rule mechanism, if that mechanism contains only very general rules to derive regular, default patterns. Such gradient effects are more likely to come from the analogical mechanism, provided that at least some regular forms are listed and can exert an analogical effect. Certainly there is nothing intrinsic in the dual-mechanism approach that precludes the listing of regular forms, and there is evidence that regulars are in fact listed in particular circumstances, such as when they fall into a predominately irregular neighborhood (Pinker & Prince 1994:331). But here, too, there is a problem: the regular forms that contribute to the effect observed here do not have any such reason to be stored. For example, the novel *-áre* form [g^warminare] could conceivably get an analogical boost from existing words like [kulminare] 'culminate', [kamminare] 'walk', [terminare] 'end', and maybe even [g^wardare] 'watch', if they were listed. These verbs are all in reliably regular neighborhoods, however, so there would hardly be any need to list them to protect them from irregularization.[16] In fact, the strongest islands of reliability for regulars are those with no irregular neighbors, and it is generally assumed that such regulars are not listed (Clahsen 1999).

The upshot is that the dual-mechanism model must be modified to account for the Italian data. This could possibly be accomplished by allowing more extensive listing of regulars; alternatively, it could be accomplished by allowing the grammar to contain a richer set of rules, as proposed here.

4.5. Other models of morphology. The finding that a regular, default process applies with greater certainty in some phonological environments than in others is a problem not only for the dual-mechanism model, but also for many other traditional approaches to morphology. Advocates of the dual mechanism model have been most explicit about the claim that speakers do not form subgeneralizations about regular processes, but, in fact, this claim is implicit in other models as well. Most linguistic models of morphological irregularity and defaults share a similar architecture: individual words may be specified for irregularities in their inflected forms, and if they are not specified for any, then more general rules fill in the default, regular forms. This is true, for example, of Aronoff's (1976) word-based morphology, as well as distributed morphology (Halle & Marantz 1994), minimalist morphology (Wunderlich 1996), and DATR (Evans & Gazdar 1996). For all of these theories, the problem is that a nonce

[15] Although environment reliability does a better job of modeling Prasada and Pinker's English results, it remains an open question whether a model based on lexical neighborhoods could capture the Italian data. The main goal of this article is to show that there is an effect of existing regulars on novel *-áre* forms in Italian, and that at least one model can capture it. There are certainly many other ways to model lexical effects, and some of them may work as well or better in capturing this data.

[16] The *Dizionario Italiano Sabatini Coletti* (Sabatini 1998) lists only two verbs whose roots end in *-min-* that are not in the *-áre* class (*infemminire* and *inverminire*), neither of which is in current use; there are no verbs containing [g^ward] in any class other than *-áre*.

word could not possibly have any listed irregularities, and should therefore fall through automatically to the most general, default rule. In all cases, the theory could be modified to allow for a guessing component that tries to guess if the new word might possibly be irregular, but the results testing the trade-off hypothesis suggest that this alone is not sufficient. In fact, it appears that we also need a way to assign the default pattern with greater certainty when the input 'looks especially regular'.[17] In the minimal generalization model employed here, this is accomplished by allowing the regular process to be described by many different rules at all levels of generality, each with its own confidence value.

5. CONCLUSION. The results of my study can be summarized as follows: First, adult native Italian speakers find that for all verb classes some novel items sound better than others. This contradicts a number of previous findings that claim that regular morphological processes are uniformly applicable to all novel words. Second, ratings of novel words are significantly correlated with numerical environment reliability values based on the minimal generalization model of morphological learning. The fact that human ratings correlate well with neighborhood reliability values indicates that speakers have knowledge about environments and reliabilities for all inflectional classes, including the regular class. Because the ratings in this study are of novel words and not of existing ones, one must conclude that speakers are somehow able to generalize their knowledge about the distribution of classes in the existing lexicon.

The model outlined in §2 represents this knowledge as a large set of symbolic rules, at all levels of generality. There appear to be two features that contribute to the success of this model in capturing the Italian data, as well as Prasada and Pinker's English data: First, it calculates reliabilities of morphological processes within particular phonological environments, expressed as the structural description format of symbolic rules, rather than focusing on lexical neighborhoods. In §4.4, it was argued that predictions based on phonological environments found an effect for English, where a lexical neighborhoods approach failed to find one. Second, this model retains information about reliabilities in many specific environments, rather than simply keeping the single most general rule to derive regulars.

The current findings have implications for symbolic and nonsymbolic approaches to morphology. For advocates of exemplar-based and analogical approaches, the challenge is to find some definition of similarity and lexical neighborhoods that can capture the item-by-item differences that are observed for novel regulars, both in the Italian data and in Prasada and Pinker's English data. For symbolic grammar-based approaches, the challenge is to determine where local generalizations about regular morphological processes reside in the linguistic system. Under the model I have employed here, they are simply listed alongside the more general rules, in a single, probabilistic grammar that includes both regular and irregular processes. Under a more traditional model, with a limited number of very general rules, knowledge about subgeneralizations may be viewed as 'workbench material', created during the grammatical acquisition stage and retained in some form in the adult system. This view would be incomplete, however, without three additional mechanisms: (1) a procedure that can identify the 'real' rules from amidst many more effective local variants of the rules on the workbench, (2) an

[17] Note that attempts to get different defaults to apply to different words, such as Fraser and Corbett's (1997) analysis of Arapesh nouns, are not directly relevant here—in the present case, there is only one class (*-áre*) that acts as a default, so it would not help to be able to assign other classes as the default to specific subclasses of words.

explicit theory of how and where workbench knowledge is represented, and (3) an explanation of why some tasks encourage speakers to use workbench knowledge instead of the rules of their grammar. But no matter what model is used to derive productive inflections, the results presented here show that it must be more than simply a single, context-free rule.

Appendix A: test items

Items are listed in experimental order. Ratings range from 1 (worst) to 7 (best). Predicted values from the minimal generalization learner are given for each novel item in each class and range from 0 (worst) to

NO.	ROOT	TRANSCRIPTION	ROOT RATING	MEAN RATINGS				ENVIRONMENT RELIABILITY			
				-áre	**-ere**	**-ére**	**-íre**	**-áre**	**-ere**	**-ére**	**-íre**
2	*raffang-*	[raf:aŋg]	4.06	5.32	3.58	1.90	2.32	0.76	0.74	0.02	0.09
3	*zuisd-*	[dzuizd]	2.02	4.18	2.98	1.84	3.70	0.76	0.36	0.05	0.14
4	*lavess-*	[lavɛs:]	3.28	5.26	2.86	2.26	3.46	0.90	0.11	0.02	0.09
5	*boned-*	[boned]	3.34	3.82	3.34	2.62	4.54	0.76	0.36	0.15	0.24
6	*guarmin-*	[g^{w}armin]	3.88	5.56	2.32	2.44	4.00	0.96	0.11	0.05	0.12
8	*qualud-*	[k^{w}alud]	3.46	5.02	4.60	2.38	3.76	0.76	0.92	0.05	0.14
10	*muning-*	[muniŋg]	3.70	4.06	4.78	2.20	3.88	0.76	0.42	0.11	0.09
11	*pernid-*	[pernid]	3.94	5.02	4.00	2.14	3.76	0.76	0.36	0.05	0.14
12	*caffud-*	[kaf:ud]	3.40	4.66	4.06	2.50	3.88	0.76	0.59	0.05	0.14
13	*taffed-*	[taf:ed]	3.46	4.90	3.22	3.58	4.18	0.76	0.36	0.15	0.24
15	*stamign-*	[stamiɲ:]	3.64	5.62	2.80	2.38	3.64	0.94	0.11	0.02	0.09
16	*dolific-*	[dolifik]	4.66	5.92	2.62	1.90	2.20	0.99	0.10	0.02	0.09
17	*ganizz-*	[ganidz:]	4.54	5.92	2.62	2.08	3.58	0.99	0.11	0.02	0.09
19	*giabban-*	[ʤabban]	3.94	5.62	2.02	2.26	3.76	0.78	0.11	0.05	0.12
20	*rabad-*	[rabad]	3.58	5.38	2.98	2.98	5.20	0.76	0.36	0.38	0.18
22	*panizz-*	[panidz:]	4.72	6.04	2.08	1.90	3.52	0.99	0.11	0.02	0.09
23	*ligun-*	[ligun]	3.94	5.38	2.74	2.50	4.42	0.78	0.11	0.05	0.37
25	*grobm-*	[grobm]	2.08	3.64	1.84	1.90	4.18	0.76	0.14	0.02	0.09
26	*sfagon-*	[sfagon]	4.54	5.50	1.66	2.08	4.12	0.92	0.11	0.05	0.13
27	*vegheggi-*	[vegeʤ:]	4.90	6.10	2.98	1.84	3.16	0.99	0.11	0.02	0.09
28	*colond-*	[kolond]	4.12	5.20	3.82	2.02	4.24	0.76	0.54	0.05	0.14
29	*sovvrogli-*	[sov:roʎʎ]	3.16	5.38	3.94	1.84	3.22	0.78	0.48	0.02	0.09
30	*rubind-*	[rubind]	4.42	5.56	3.88	2.02	4.24	0.76	0.54	0.05	0.14
31	*barud-*	[barud]	4.36	5.56	4.30	1.96	4.72	0.76	0.59	0.05	0.14
32	*spinad-*	[spinad]	3.64	4.66	3.46	2.68	3.82	0.76	0.36	0.31	0.18
33	*blins-*	[blins]	2.68	4.72	2.92	1.96	3.82	0.87	0.11	0.02	0.09
34	*funad-*	[funad]	3.88	4.42	3.88	2.44	3.94	0.76	0.36	0.31	0.18
36	*cabet-*	[kabet]	3.40	4.72	3.34	2.26	3.58	0.79	0.16	0.02	0.09
37	*mastand-*	[mastand]	3.28	5.14	3.46	2.20	4.00	0.76	0.54	0.05	0.16
38	*mussid-*	[mus:id]	3.40	4.42	4.24	2.14	3.70	0.76	0.36	0.05	0.14
39	*braips-*	[braips]	2.50	4.60	2.98	2.20	3.88	0.87	0.11	0.02	0.09
40	*scincad-*	[ʃɪŋkad]	3.16	3.88	2.86	3.10	3.28	0.76	0.36	0.83	0.18
41	*emmett-*	[em:et:]	4.36	3.82	5.68	1.96	3.40	0.79	0.95	0.02	0.09
42	*fraquin-*	[frak^{w}in]	3.58	5.32	2.98	1.84	3.52	0.92	0.11	0.05	0.12
43	*aggiend-*	[aʤ:ɛnd]	3.28	4.84	4.12	1.90	3.76	0.76	0.84	0.05	0.09
44	*genic-*	[ʤenik]	4.30	4.90	2.26	1.66	2.92	0.99	0.10	0.02	0.09
46	*pasced-*	[paʃ:ed]	4.12	3.94	4.96	2.74	3.16	0.76	0.36	0.15	0.24
47	*svegnod-*	[zveɲod]	3.10	4.48	2.74	2.38	3.34	0.76	0.36	0.05	0.14
48	*ghirilb-*	[girilb]	3.16	4.18	3.82	2.32	4.00	0.76	0.11	0.02	0.15
50	*pebcr-*	[pebkr]	1.96	3.76	2.44	1.78	2.68	0.76	0.11	0.02	0.13
51	*bragod-*	[bragod]	4.00	5.32	3.16	3.40	4.12	0.76	0.36	0.54	0.14
52	*addung-*	[ad:uŋg]	4.36	4.06	5.26	2.02	3.34	0.76	0.74	0.02	0.09
53	*famid-*	[famid]	3.52	5.02	3.70	2.14	3.58	0.76	0.36	0.05	0.14
54	*ravelod-*	[ravelod]	3.22	4.48	3.34	2.44	4.00	0.76	0.36	0.05	0.14

Table A.1. Test items, with mean ratings and predicted values.

1 (best). In addition to the nonce words listed in the table, 10 real Italian words were also presented for rating. These were: *implor-* 'implore' (item 1), *boll-* 'stamp' (*-áre*)/'boil' (*-íre*) (7), *tratten-* 'hold back' (9), *commerci-* 'deal (in)' (14), *smuov-* 'move' (18), *consegn-* 'deliver' (21), *sfuggi-* 'escape' (24), *plasm-* 'mold' (35), *fulmin-* 'strike by lightning' (45), and *disperd-* 'scatter' (49).

Appendix B: phonological environments for -are.

The tables below list a few of the most reliable phonological environments for *-áre* suffixation, as calculated by the Albright and Hayes computational model. Adjusted reliability values are the lower confidence limit (α = .75) of the raw reliability, which equals the hits divided by the scope. A full list of all phonological environments explored for all conjugation classes, along with their reliability values, can be downloaded from http://ling.ucsc.edu/~albright/papers/italian.html.

RELIABLE ENVIRONMENTS (= islands of reliability)	HITS	SCOPE	ADJUSTED RELIABILITY (α = .75)
-o → *-are* / X [idz:]	65	65	0.985
-o → *-are* / X [ts]	104	105	0.978
-o → *-are* / X [it]	43	43	0.978
-o → *-are* / X [ts:]	90	91	0.975
-o → *-are* / X [ʤ:]	33	33	0.971
-o → *-are* / X [fik]	32	32	0.970
-o → *-are* / X [ifik]	31	31	0.969
-o → *-are* / X [lidz:]	30	30	0.968
-o → *-are* / X [ɲɲ]	24	24	0.961
-o → *-are* / X [eʤ:]	23	23	0.959
-o → *-are* / X [alidz:]	20	20	0.953
-o → *-are* / X [eg]	19	19	0.950
-o → *-are* / X [min]	18	18	0.948
-o → *-are* / X [jon]	18	18	0.948
-o → *-are* / X [tsj]	16	16	0.941
-o → *-are* / X [kj]	16	16	0.941
-o → *-are* / X [pp]	15	15	0.938
-o → *-are* / X [ass]	14	14	0.934
-o → *-are* / X [ss]	31	32	0.930
-o → *-are* / X [ir]	13	13	0.928
-o → *-are* / X [eɲɲ]	13	13	0.928
-o → *-are* / X [tsjon]	13	13	0.928
-o → *-are* / X [ʃ]	13	13	0.928
-o → *-are* / X [ridz:]	12	12	0.923
-o → *-are* / X [el]	12	12	0.923
-o → *-are* / X [tifik]	12	12	0.923
-o → *-are* / X [tr]	27	28	0.920
-o → *-are* / X [kk]	27	28	0.920
-o → *-are* / X [ont]	11	11	0.916
-o → *-are* / X [ok]	11	11	0.916
-o → *-are* / X [ntsj]	11	11	0.916
-o → *-are* / X [port]	10	10	0.908
-o → *-are* / X [ntr]	10	10	0.908
-o → *-are* / X [iʎʎ]	10	10	0.908
-o → *-are* / X [kkj]	10	10	0.908
-o → *-are* / X [ol]	34	36	0.904
-o → *-are* / X [s]	44	47	0.901
-o → *-are* / X [ʤ]	44	47	0.901
-o → *-are* / X [okk]	9	9	0.898

Table B.1. Some reliable environments for *-áre*.

RELIABLE ENVIRONMENTS (= islands of reliability)	HITS	SCOPE	ADJUSTED RELIABILITY (α = .75)
-o → *-are* / X [str]	9	9	0.898
-o → *-are* / X [vj]	9	9	0.898
-o → *-are* / X [ess]	9	9	0.898
-o → *-are* / X [tidz:]	9	9	0.898
-o → *-are* / X [og]	9	9	0.898
-o → *-are* / X [ets:]	9	9	0.898
-o → *-are* / X [l]	113	124	0.891
-o → *-are* / X [ment]	19	20	0.889

TABLE B.1. (continued) Some reliable environments for *-áre.*

REFERENCES

ALBRIGHT, ADAM. 2002a. The identification of bases in morphological paradigms. Los Angeles: University of California, Los Angeles dissertation.

ALBRIGHT, ADAM. 2002b. The lexical bases of morphological well-formedness. Morphology 2000: Selected papers from the 9th morphology meeting, Vienna, 25–27 February 2000, ed. by S. Bendjaballah, W. U. Dressler, O. E. Pfeiffer, and M. D. Voeikova, 5–15. Amsterdam: John Benjamins.

ALBRIGHT, ADAM, and BRUCE HAYES. 2002. Modeling English past tense intuitions with minimal generalization. Proceedings of the ACL 2002 Workshop on Morphological and Phonological Learning. ACL Publications.

ARONOFF, MARK. 1976. Word formation in generative grammar. Cambridge, MA: MIT Press.

BERENT, IRIS; JOSEPH SHIMRON; and STEVEN PINKER. 1999. Default nominal inflection in Hebrew: Evidence for mental variables. Cognition 72.1.1–44.

BERKO, JEAN. 1958. The child's learning of English morphology. Word 14.2–3.150–77.

BYBEE, JOAN. 1995. Regular morphology and the lexicon. Language and Cognitive Processes 10.5.425–55.

BYBEE, JOAN, and CAROL MODER. 1983. Morphological classes as natural categories. Language 59.251–70.

BYBEE, JOAN, and DAN SLOBIN. 1982. Rules and schemas in the development and use of the English past tense. Language 58.265–89.

CLAHSEN, HARALD. 1999. Lexical entries and rules of language: A multidisciplinary study of German inflection. Behavioral and Brain Sciences 22.991–1060.

CLAHSEN, HARALD, and MONIKA ROTHWEILER. 1992. Inflectional rules in children's grammars: Evidence from the development of participles in German. Yearbook of Morphology 1992. 1–34.

COLTHEART, MAX; EILEEN DAVELAAR; JON TORFI JONASSON; and DEREK BESNER. 1977. Access to the internal lexicon. Attention and performance 6, ed. by Stanislav Dornic, 535–55. Hillsdale, NJ: Erlbaum.

DARDANO, MAURIZIO, and PIETRO TRIFONE. 1985. La Lingua Italiana. Bologna: Zanichelli.

DAVIS, STUART, and DONNA JO NAPOLI. 1994. A prosodic template in historical change: The passage of the Latin second conjugation into Romance. Torino: Rosenberg & Sellier.

DE MAURO, TULLIO; FEDERICO MANCINI; MASSIMO VEDOVELLI; and MIRIAM VOGHERA. 1993. Lessico di frequenza dell'italiano parlato. Milano: Etaslibri.

DIXON, ROBERT M. W. 1977. A grammar of Yidiɲ. Cambridge: Cambridge University Press.

EDDINGTON, DAVID. 2002. Dissociation in Italian conjugations: A single-route account. Brain and Language 81.291–302.

EVANS, ROGER, and GERALD GAZDAR. 1996. DATR: A language for lexical knowledge representation. Computational Linguistics 22.2.167–216.

FRASER, NORMAN, and GREVILLE CORBETT. 1997. Defaults in Arapesh. Lingua 103.25–57.

FUJIWARA, MIHO, and MICHAEL T. ULLMAN. 1999. The computation of default suffixation in Japanese adjectival past tense forms. Paper presented at the 12th annual CUNY conference on human sentence processing, New York, March 1999.

HALLE, MORRIS, and ALEC MARANTZ. 1994. Some key features of distributed morphology. MIT Working Papers in Linguistics 21.275–88.

HARE, MARY, and JEFFREY ELMAN. 1995. Learning and morphological change. Cognition 56.61–98.

KÖPCKE, KLAUS-MICHAEL. 1993. Schemata bei der Pluralbildung im Deutschen: Versuch einer kognitiven Morphologie, vol. 47: Studien zur deutschen Grammatik. Tübingen: Gunter Narr Verlag.

LANDAUER, THOMAS K., and LYNN A. STREETER. 1973. Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. Journal of Verbal Learning and Verbal Behavior 12.119–31.

LUCE, PAUL A. 1986. Neighborhoods of words in the mental lexicon. Technical Report 6, Speech Research Laboratory, Department of Psychology, Indiana University.

MACWHINNEY, BRIAN, and JARED LEINBACH. 1991. Implementations are not conceptualizations: Revising the verb learning model. Cognition 40.121–57.

MAIDEN, MARTIN. 1995. Review of A prosodic template in historical change: The passage of the Latin second conjugation into Romance, by Stuart Davis and Donna Jo Napoli. Forum Italicum 29.391–93.

MARCUS, GARY; URSULA BRINKMANN; HARALD CLAHSEN; RICHARD WIESE; and STEVEN PINKER. 1995. German inflection: The exception that proves the rule. Cognitive Psychology 29.186–256.

MAROTTA, GIOVANNA. 1997. Review of A prosodic template in historical change: The passage of the Latin second conjugation into Romance, by Stuart Davis and Donna Jo Napoli. Studi Italiani di Linguistica Teorica e Applicata 26.595–600.

MIKHEEV, ANDREI. 1997. Automatic rule induction for unknown-word guessing. Computational Linguistics 23.3.405–23.

NAKISA, RAMIN CHARLES; KIM PLUNKETT; and ULRIKE HAHN. 1997. Single and dual-route models of inflectional morphology. Cognitive models of language acquisition, ed. by Peter Broeder and Jaap Murre, 201–22. Cambridge, MA: MIT Press.

NAPOLI, DONNA JO, and IRENE VOGEL. 1990. The conjugations of Italian. Italica 67.4.479–502.

NEWMAN, ROCHELLE; JAMES SAWUSCH; and PAUL LUCE. 1997. Lexical neighborhood effects in phonetic processing. Journal of Experimental Psychology: Human Perception and Performance 23.3.873–89.

ORSOLINI, MARGHERITA; RACHELE FANARI; and HUGO BOWLES. 1998. Acquiring regular and irregular inflection in a language with verb classes. Language and Cognitive Processes 13.4.425–64.

ORSOLINI, MARGHERITA, and WILLIAM MARSLEN-WILSON. 1997. Universals in morphological representation: Evidence from Italian. Language and Cognitive Processes 12.1.1–47.

PINKER, STEVEN. 1999. Words and rules: The ingredients of language. New York: Basic Books.

PINKER, STEVEN, and ALAN PRINCE. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. Cognition 28.73–193.

PINKER, STEVEN, and ALAN PRINCE. 1994. Regular and irregular morphology and the psychological status of rules of grammar. The reality of linguistic rules, ed. by Susan D. Lima, Roberta L. Corrigan, and Gregory K. Iverson, 321–51. Amsterdam: John Benjamins.

PRASADA, SANDEEP, and STEVEN PINKER. 1993. Generalization of regular and irregular morphological patterns. Language and Cognitive Processes 8.1–56.

RUMELHART, DAVID E., and JAMES L. MCCLELLAND. 1986. On learning the past tenses of English verbs. Parallel distributed processing: Explorations in the microstructure of cognition, ed. by James L. McClelland, David E. Rumelhart, and the PDP Research Group, vol. 2: Psychological and biological models, 216–71. Cambridge, MA: Bradford Books/MIT Press.

SABATINI, FRANCESCO. 1998. Dizionario Italiano Sabatini Coletti. Giunti Multimedia.

SAY, TESSA. 1998. Regular and irregular inflection in the mental lexicon: evidence from Italian. Essex Graduate Student Papers in Language and Linguistics 2.55–70.

SAY, TESSA, and HARALD CLAHSEN. 2000. Words, rules and stems in the Italian mental lexicon. Storage and computation in the language faculty, ed. by Sieb Nooteboom, Fred Weerman, and Frank Wijnen, 93–129. Dordrecht: Kluwer.

SKOUSEN, ROYAL. 1989. Analogical modeling of language. Dordrecht: Kluwer.
ULLMAN, MICHAEL T. 1999. Acceptability ratings of regular and irregular past-tense forms: Evidence for a dual-system model of language from word frequency and phonological neighborhood effects. Language and Cognitive Processes 14.1.47–67.
ULLMAN, MICHAEL T. 2001. The declarative/procedural model of lexicon and grammar. Journal of Psycholinguistic Research 30.1.37–69.
WRIGHT, ROGER. 1997. Review of A prosodic template in historical change: The passage of the Latin second conjugation into Romance, by Stuart Davis and Donna Jo Napoli. Studies in Language 21.169–73.
WUNDERLICH, DIETER. 1996. Minimalist morphology: The role of paradigms. Yearbook of morphology 1995. 93–114.

Dept. of Linguistics
University of California, Santa Cruz
1156 High Street
Santa Cruz, CA 95064-1077
[albright@ling.ucsc.edu]

[Received 4 December 2000;
revision received 9 January 2002;
accepted 7 June 2002]