Modeling Morphological Subgeneralizations[*]

Claire Moore-Cantwell and Robert Staubs
University of Massachusetts Amherst

1 Introduction

Exceptions to morphological regularities often pattern together phonologically. In the English past tense system, exceptions to the regular 'Add /-d/' rule frequently inhabit 'Islands of Reliability' (Albright & Hayes, 2003), in which a group of words take the same irregular past and also pattern together on a set of phonological characteristics. For example *ring*, *sing*, and *stink* are part of a group of words which all undergo a particular vowel change ($\text{ɪ} \rightarrow \text{æ}$) to realize the past tense, and all tend to share certain phonological characteristics, such as ending in a velar, ending in a nasal, or beginning with an s-stop cluster.

In a system like the English past, the behavior of each individual lexical item must be memorized, but adults nonetheless seem to have active implicit knowledge of both the overall pattern (the regular past) and the 'subgeneralizations' ($\text{ɪ} \rightarrow \text{æ}/ \_ \text{[+ velar]}$). Bybee & Moder (1983), Prasada & Pinker (1993), and Albright & Hayes (2003) all conducted productivity tests of the English past tense system, finding that although English speakers are most likely to choose a regular past for a novel verb, they also occasionally produce irregulars (e.g. spling $\rightarrow$ splang). When they do, they respect the subgeneralizations, choosing the irregular past based on the behavior of words which are phonologically similar to the base.

We model this implicit knowledge of morphological subgeneralizations through the interaction of a structured lexicon and a Maximum Entropy grammar. Words that pattern together with respect to a particular morphological process (e.g. all words that express the past by changing [ɪ] to [æ]) are grouped together into a 'bundle', which is indexed to a constraint expressing the change that these words must undergo to realize the morpheme. These morphological 'operational constraints' compete with markedness and faithfulness constraints in the phonological component of the model. Each bundle has its own phonological representation which is the average of the representations of the lexical items in it, and a novel lexical item can choose which bundle to belong to based on its similarity to each bundle's representation.

2 Background

A system with a single regular rule and a series of memorized exceptions cannot capture subregularities like the 'ɪ $\rightarrow$ æ' generalization. Models such as Rumelhart & McClelland (1986) represent systems with subregularities as a connectionist network in which there are no explicit rules. Rather, connections between features of the present and past are strengthened through exposure to existing present-past pairs. Subregularities are represented in the same way as the regular past generalization—as a pattern of connection weights between features of the present and features of the past.

Similarly, analogical models such as e.g. Nosofsky (1990) determine a novel past for both regulars and irregulars based on the base's phonological similarity to an array of existing words. Connectionist and analogical models have been criticized on the grounds that they don't have a way to account for the generality of the phonology-morphology interaction (e.g. Pinker & Prince, 1988). In English, the –t/–d/–əd alternation of the past shares conditioning factors with the –s/–z/–əz alternation of the plural and possessive markers: all three morphemes agree in voicing with the final segment of their base, and avoid repeated alveolar or sibilant sounds by schwa epenthesis. In connectionist and analogical models, there is no explicit statement of these abstract conditioning factors. Instead, the generality of the voicing assimilation and schwa epenthesis

Proceedings of *Phonology 2013*
Completed February 18, 2014

processes is treated as coincidental.

Dual-route models like Pinker & Prince (1988); Pinker (1999); Marcus et al. (1995); Clahsen et al. (2003) capture both subregularities within irregulars and general phonological processes like voicing agreement by dividing up the system into two parts: an analogical mechanism which operates on the lexicon, and a rule-based grammar which applies the regular rule and phonology. This division of labor predicts that factors such as lexical frequency and similarity between forms should affect speakers' choice of irregular forms for a novel verb, but not their choice of the regular. That is, a novel verb which is phonologically similar to some members of a class of irregulars should be more likely to behave like that class than a word which is not similar to it. However, a novel verb which is not phonologically similar to any particular regular should be just as likely to take a regular past as a verb which is very similar to a particular regular, or group of regulars. Prasada & Pinker (1993) find that novel verbs which are phonologically similar to irregulars in a particular group pattern with those irregulars more often than verbs which are dissimilar. They find, on the other hand, that novel verbs which are dissimilar to any regular (they use words that are phonotactically rare in English, e.g. *ploamph*) are just as likely to take a regular past as novel verbs which are similar to regulars.

Albright & Hayes (2003) term phonological contexts in which a particular type of past tense change is especially likely to occur 'Islands of Reliability' (IOR's). Contra Prasada & Pinker (1993), they find that these Islands of Reliability are respected for regulars as well as for irregulars. For example, verbs which end in _ [f] always take a regular past, and novel verbs which end in _ [f] also receive a regular past more often than verbs which are not in an irregular IOR but which do not end in _ [f]. Based on this finding, they argue that a single mechanism must be responsible for speakers' knowledge of both regular and irregular generalizations. They propose a rules-only model, the Minimal Generalization Learner (MGL), which learns rules at many levels of generality, starting from individual lexical items.

To illustrate how the MGL learns its rules, consider the case of the regular rule, in Figure 1. The learner starts with a rule specific to each lexical entry, here *shine* and *consign*. Then, the learner can generalize across just those two lexical items with the same change, creating a new rule that is minimally general. Eventually more and more generalization will occur so that the maximally general form of the rule emerges. All the rules, not just the most general ones, stay in the system and help to determine the choice of past for novel items. A rule's *scope*, the number of forms it applies to, and *confidence*, the percentage of time it produces the correct output on trained lexical items, determine how likely it is to be used.

$$\emptyset \rightarrow d \ / \ [\qquad\qquad \int \qquad \text{ain} \quad \underline{\quad} \ ]_{[+ \text{ past}]}$$
$$\emptyset \rightarrow d \ / \ [\quad \text{k\ə n} \quad \text{s} \qquad \text{ain} \quad \underline{\quad} \ ]_{[+ \text{ past}]}$$
$$\Rightarrow \quad \emptyset \rightarrow d \ / \ [\quad \text{X} \quad [\text{vcls}\,] \quad \text{ain} \quad \underline{\quad} \ ]_{[+ \text{ past}]}$$
$$\vdots$$
$$\Rightarrow \quad \emptyset \rightarrow d \ / \ [\qquad\qquad \text{X} \qquad\quad \underline{\quad} \ ]_{[+ \text{ past}]}$$

Figure 1: Example of rule generalization in MGL from words like *shine* and *consign*.

The MGL predicts that for any novel verb, speakers should overwhelmingly prefer the regular past tense. This is because the regular rule in Figure 1, although it has exceptions, has a much wider scope than any irregular rule, and reasonably high confidence. Within the space of regulars, there are rules which are not the most general rule, but which have very high, or even 100% confidence, e.g. $\emptyset \rightarrow d \ / \ [\text{X f} \ \underline{\quad} \ ]_{[+ \text{ past}]}$. Irregular IOR's are represented as similar subgeneralizations: $[\text{ɪ}] \rightarrow [\text{ʌ}] \ / \ [\text{X}\{\text{l,r}\} \ \underline{\quad} \ \text{ŋ}]_{[+ \text{ past}]}$, for e.g. *spring, cling*. Novel words within an irregular IOR will be more likely to take that irregular than novel words not in that IOR. Also, novel words within a regular IOR will be more likely to take a regular than words in neither an irregular nor regular IOR.

The MGL runs into a similar problem as purely analogical models however, namely that it does not straightforwardly capture the interaction of exceptionless phonology and exceptionful morphology (Albright and Hayes section 2.3.3). Consider the most general rule in Figure 1: this rule will add [d] to any verb, creating in some cases phonotactically ill-formed outputs like [nidd] for *need*. Even if the system learns a subgeneralization like $\emptyset \rightarrow \text{əd} \ / \ [\text{X} \ \{\text{t,d}\} \ \underline{\quad} \ ]_{[+ \text{ past}]}$, the more general rule will sometimes apply. Albright and Hayes employ a constraint-based 'phonological filter', which examines the output of the morphological rules, and removes phonotactically illegal outputs from consideration. Without this filter, the MGL produces the –t/–d/–əd alternation as an exceptional generalization similar to the irregular generalizations.

The MGL with a phonological filter requires that phonology be learned entirely before morphology, and does not allow the acquisition of the two systems to interact. The system cannot 'choose' whether to attribute patterned variation in the surface form of a morpheme to the action of morphology or phonology. Hayes (2004) discusses in detail the claim that phonotactics are learned entirely before morphology, finding that it is mostly plausible, but that in some cases the learner needs access to morphological information to discover a phonological constraint. He proposes a 'backtracking' mechanism to fix these early errors.

We propose instead that phonology and morphology can be learned at least somewhat simultaneously. In our model, the –t/–d/–əd alternation can be represented as a morphological generalization, in which words that end in voiced sounds are associated with an 'ADD –/d/' rule, words that end in coronal stops are associated with an 'ADD –/əd/' rule, and words that end in voiceless sounds are associated with an 'ADD –/t/ rule. In this case, the associations are coincidental: no connection is drawn between the rule applied and the type of words it applies to. However, our model prefers to represent this alternation phonologically, with a single regular class, associated with a rule like 'ADD –/d/'. The output of this rule is then altered by the phonological component which contains constraints like '*[+ voice][− voice]'. The model typically learns phonological conditioning for this alternation precisely because the factors conditioning the alternation are not coincidental. The model is capable of inducing relatively simple phonological constraints which do the same job as the division into different morphological categories would, and because the phonological constraints correctly produce the alternations categorically, the phonological approach better matched the training data.

3   An integrated model of the lexicon and morphology

3.1   *Operational constraints*   In our model, lexical items are grouped together into *bundles*. These bundles are the units of morphological generalization, acting as intermediate abstractions between individual items and overall patterns. Each lexical item is indexed to a set (possibly null) of corresponding bundles. Each bundle is affiliated with an *operational constraint*.

Operational constraints motivate the use of a particular operation for the exponence of a particular morphosyntactic feature. These constraints are violated when a lexical item evaluated in a particular bundle does not use the corresponding operation noted in its operational constraint. These operations apply to morphological bases, generating one underlying representation (UR) from another. Phonological operations may follow morphological ones, so the resulting surface representations (SR) may not obviously reflect the morphological changes.

(1) $\Phi(\text{BASE})_i$: X → Y:
Candidates are triples of a base, underlying representation, and surface representation (B, UR, SR). Assign a violation for every triple such that B is an instance of BASE, UR contains an exponent of the morphosyntactic feature Φ, the mapping is evaluated with bundle i, and the mapping B → UR does not contain the mapping X → Y.

For example, a constraint for the past tense of *sing* (assuming it is in a bundle indexed as 2) might look like the following:

(2)   PAST(PRESENT)$_2$: ɪ → æ
≈ 'The past tense of bundle 2 words should be formed from the present tense by changing ɪ to æ.'

This constraint schema has several parts, not all of which were varied in this study. Φ is intended to represent any particular morphosyntactic feature requiring exponence, but we only study the past tense of verbs here. The operational mapping $X \rightarrow Y$ is varied to represent different morphological generalizations. The bundle index t is not at present significant to any larger theory of lexical structure beyond the workings of these constraints. Finally, in the present paper, BASE is always assumed to be the first person singular of the verb, denoted here by PRESENT. The past tense is assumed to be derived from this other form of the verb, which is taken as basic. The mapping between the base form and its resulting past tense is given, not discovered. In future work, we plan to relax assumptions about BASE, allowing for base discovery or operation-specific bases.

These constraints interact with more traditional markedness and faithfulness constraints, thus providing for the interaction of phonology and morphology. The candidates are evaluated by a Maximum Entropy

(Goldwater & Johnson, 2003) grammar: constraint weights are used to derive candidate harmonies (weighted sums of constraint violations) and candidate probabilities are proportional to the exponential of harmony. We refer to the output of this probabilistic choice as the *optimum*, though it may not truly be optimal for the weights.

(3) $p(x) \propto e^{H_x}$

The parts of an optimization are exemplified in Figure 2. NEED has a bundle index 1, and only this, so that bundle is used in the optimization. Underlying representations are generated according to the available bundles, adding /d/ and changing /i/ to /ɛ/. The surface-level candidates are formed from these representations by a phonological GEN. The only morphological constraint that is relevant is the one for the regular—ADD $-$/d/$_1$, because this is the only one indexed to bundle 1. The other morphological constraints are essentially inactive. The optimization chooses the correct output by simultaneously using this single morphological constraint and the phonological constraints.

Operational constraints are related to a number of predecessors in Optimality Theory-like theories. Their operational nature essentially involves compelling a sort of unfaithfulness between the base and the feature UR, so they are related to anti-faithfulness as described by Alderete (2001). However, operational constraints differ in the specific nature of this unfaithfulness. Realizational constraints (e.g. Aronoff & Xu, 2010) offer another parallel. These constraints motivate particular forms for outputs. However, they make no statement about *changes* and are necessarily surface-true.

UR constraints (Boersma, 2001) motivate a choice in UR, as do operational constraints, but this choice is not in principle related to the base. This distinction is closely tied to the 'operational' character of the model. This aspect of our approach is related to work with targeted constraints (Wilson, 2013), and also to other work in Harmonic Serialism (Prince & Smolensky, 1993; McCarthy, 2000) and Optimality Theory with Candidate Chains (McCarthy, 2007) such as Wolf's Optimal Interleaving (Wolf, 2008) and Staubs' fully operational HS version (Staubs, 2011) of OI.

This model departs from the MGL in two primary ways. First, the phonotactics of the language (English, in our example) is learned along with its morphology. This arises naturally due to the simultaneous use of markedness, faithfulness, and operational constraints. The present model therefore does not in principle require a 'phonotactic filter' or other additional mechanism for dealing with the productive phonotactics of a language. Second, the context of a 'rule' is divorced from its application. Unlike MGL, the phonological context of an operation's left-hand side is not fixed by the constraint itself. Instead, this context arises from properties of the bundle. This opens the model up to expansions: basing bundle assignment on many factors, as perhaps in lexical strata; forming bundles based on non-phonological information, as for example noun and verb stress in English.
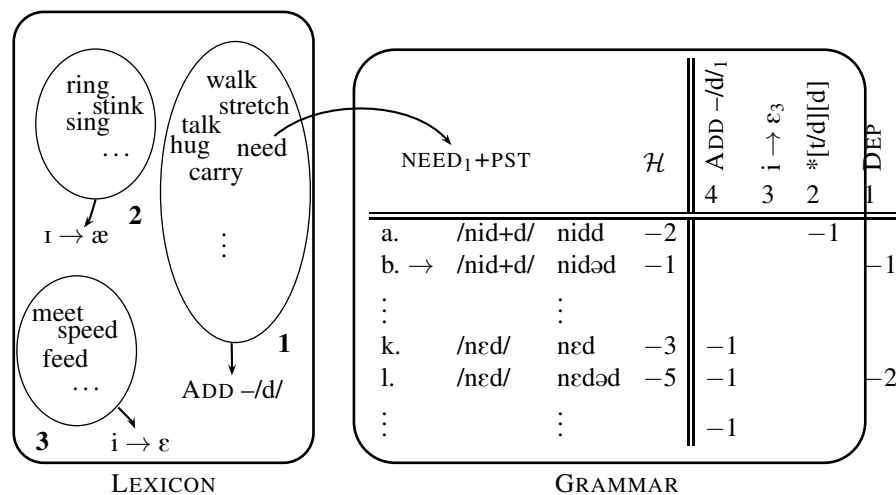

Figure 2: The meaning NEED is evaluated with respect to bundle 1 (the productive bundle). The operational constraint on this bundle assigns violations due to the shared index, others are not. The constraints *[t/d][d] and DEP together describe epenthesis for the chosen operation of adding /d/.

3.2 *Generating outputs*   To generate any output, a lexical item must be assigned to at least one bundle. This will not be true for novel items (either in training or as wugs for testing)—these will need to be assigned a bundle. Assignment to a bundle is based on similarity, as described below. Once an item has at least one bundle index, an output can be determined. First, one of the bundles is randomly chosen as the bundle for optimization. The existing operational constraints for the corresponding morphosyntactic feature provide a set of possible morphological operations to create URs. For each operation, one UR is generated from the base.

The set of URs is not enough, SRs must be generated as well in order to model phonological effects. A set of SRs is generated from each UR by the application of operations in GEN. The two types of phonological operations we consider here are feature-changing (notably, voicing-changing) and epenthesis ('schwa' epenthesis).

With the generated URs and SRs, we have full triples (B, UR, SR) as required by the constraint definition in (1). Operational constraint violations are assigned on the basis of this triple and the chosen bundle, markedness is assigned based on the SR, and faithfulness is assigned based on the operations mapping UR to SR. This process is shown in Figure 3.
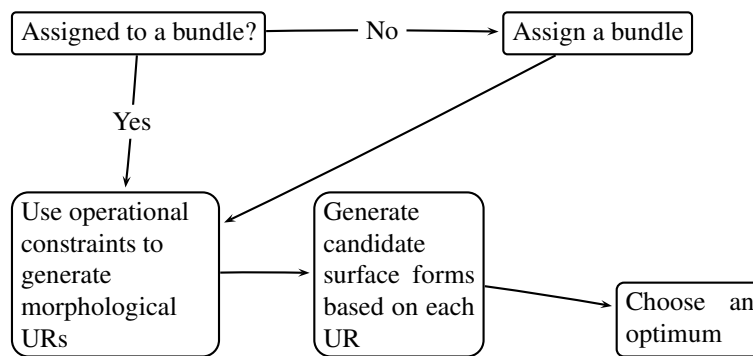

Figure 3: Schematic flow chart of how a meaning gets mapped to some output.

3.3 *Learning in the model*   Bundles need to come from somewhere. The first thing a bundle requires is some operational constraint. To get these, we align the base with the past tense (etc.), as shown in Figure 4a.[1] This alignment gives a list of different operations which map one form into the other, serving as the contents of an operational constraint. This is also how these constraints are evaluated: an alignment is performed and its operations should correspond maximally with the operational constraint.
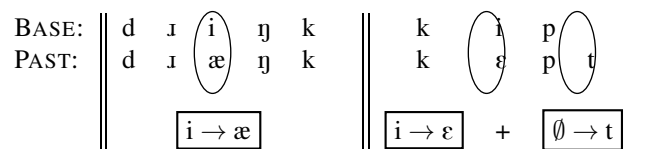
(a) A lexical item in BASE is aligned with one in PAST. The operations transforming one into the other are used in the corresponding operational constraint.
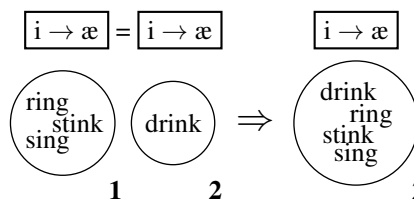
(b) Two bundles with identical operational constraints can be merged into one.

Figure 4: Examples of aligning present and past tense forms and merging their bundles. These are the typical steps on first encountering a lexical item.

---

[1]   We use Python's (2.6.6) `difflib`, related to the algorithm of Ratcliff et al. (1988). This choice is not crucial.

When a novel form needs to be assigned to a bundle, it is done on the basis of *similarity*. Phonological similarity is computed over markedness violations of the input (à la Golston, 1996). Every time a lexical item is used within a particular bundle, that bundle's mean markedness is updated by the markedness contributed by that item. This means that, in general, 'diffuse' bundles will have means close to the markedness mean of all lexical items, while more specific bundles will have means characteristic of the violation patterns of the bases to which they apply.

The violations of an item are compared with those of the mean of each bundle, yielding *distances*. These are then used to compute probabilities by exponentiating the distance and normalizing it. Thus items which are close to a bundle's mean will be likely to be assigned to it and ones farther away will be less likely.

$$(4) \quad p(i \text{ in bundle } b) = \frac{e^{-\lambda \sum_{\text{Con}} (v_i - v_b)^2}}{\sum_{\beta \in \text{Bundles}} e^{-\lambda \sum_{\text{Con}} (v_i - v_\beta)^2}}$$

where λ is a scaling parameter.

		*[ɪŋ]	*[coronal]#	*i[coronal]	*[+ voice][− voice]	[+ nasal]
	need		1	1		1
	walk				1	
regulars	ding	1				1
	mean	**0.33**	**0.33**	**0.33**	**0.33**	**0.67**
	feed		1	1		
i → ɛ	meet		1	1	1	1
	mean		**1.0**	**1.0**	**0.50**	**0.50**
	sing	1				1
i → æ	ring	1				1
	mean	**1.00**				**1.0**

Figure 5: Selected means markedness for very small bundles. Zero values are not shown. The members of each of these three bundles are scored on markedness. The mean markedness of the bundle is the mean of its members' markedness.

Figure 5 shows a small example of how the mean markedness representation works. These three small bundles have aggregate vectors consisting of their mean violations. These mean violations can typify a bundle. For example, the mean values for constraints like *i[coronal] are quite distinct for the second bundle. Novel words would be more likely to use this morphology if they scored similarly on this constraint. More concretely, this says that words like *steet* are close to the bundle mean and are likely to have past tenses like *stet*. A similar situation holds for the third bundle and the constraint against [ɪŋ].

	*[ɪŋ]	*[coronal]#	*i[coronal]	*[+ voice][− voice]	[+ nasal]
regulars	0.01	0.26	0.01	0.85	0.47
i → ɛ		1.00	1.00	0.64	0.21
i → æ	0.58	0.18		0.42	0.83

Figure 6: Markedness means across words showing a type of operation. Constraints chosen to illustrate distinctions between mean patterns of violation. These groupings are not necessarily bundles—they include potentially dissimilar contexts.

Figure 6 shows means for larger lexical groupings. Zero values are shown as empty cells. We can see that in the data a change from [i] to [ɛ] is typified by [i] followed by a coronal, while the change from [ɪ] to [æ] is associated with [ɪŋ].

The notion of similarity replaces a context in a framework based on rules alone. The change represented by a rule is contained within its operational constraint, while the rule's context is modeled as similarity to the bundle mean markedness.

In learning, present/past pairs are sampled from the lexicon. For each pair, the learner generates an optimum by the above procedure. If the output matches the sampled past tense, no change is necessary except for an update of the mean violations. If there is a mismatch, several types of change can happen. The first is that the constraints themselves are updated. This occurs on any failure and is done according to the delta rule (or SGA for MaxEnt grammar; Jäger, 2007)—the constraint weights are updated by the scaled difference between the learner's chosen form's violations and those of the target form.

(5) $w_{t+1} = w_t + \eta(v_i - v^*)$
        where v_i are the violations of the learner's chosen form and v^* are the violations of the target form.

The other two changes we consider in learning are performed only probabilistically. First, with probability p_n, the learner induces new n-gram markedness constraints from the base of the pair. Every n-gram of a particular size is added to the constraint set if it is not there already. One important purpose to this is to allow the learner to capture generalizations over the shape of bases in the future: an error in bundle assignment leads to adding more detail to bundle descriptions, hopefully leading to better assignment later on.

Second, with probability p_m, groups merge. In merger, a bundle is chosen to merge with based on the above notion of similarity. All the members of the merged bundle are now members of the new bundle, and mean markedness violation vectors are updated in line with this combination (Figure 4b). The operational constraint of the larger bundle is preserved. Merger allows bundles using similar or identical operational constraints to share effort in learning and production.

In the simulations presented here, $p_n = 0.01$ and $p_m = 0.50$. The maximum n for n-grams is set at 2: only unigrams and bigrams (sequences of one or two phones) are considered.

4   Testing the model's performance

We show in this section that, like the MGL, our model respects Islands of Reliability for both regulars and irregulars when generating novel forms. Because a single mechanism is used for both types of past, where assignment of a novel form to a bundle is always based on that form's phonological shape, regulars and irregulars are both chosen based on phonological shape. Furthermore, we demonstrate that the structure of our model, in which morphological UR's can be permuted by phonological operations, allows morphological generalizations to interface smoothly with phonological ones. During learning, the model must choose whether to attribute different realizations of a morpheme to different underlying forms (created by different operational constraints), or whether to attribute it to the action of phonological constraints.

We trained our model on 4280 present-past pairs gathered from CELEX (Baayen et al., 1993), whose lemma frequency was greater than ten. We tested the model's capability to generalize using 40 nonce verbs which Albright and Hayes created and tested with human participants.[2] These verbs are divided equally into (1) verbs in islands of reliability for irregulars (e.g. *spling*) (2) verbs in islands of reliability for regulars (e.g. *blafe*), (3) verbs in no island of reliability (e.g. *shilk*), and (4) verbs in both an irregular and a regular island of reliability (e.g. *rife*). Because the model's behavior differs from run to run, we ran the model 10 times with the same set of parameters: In all cases, the learner ran for 30 training epochs with the learning rate set to 1, and 1000 test trials per wug word were used.

Overall, the model's accuracy on trained forms was high. It correctly produced 93%–99% of trained regulars, and 69%–99% of trained irregulars, depending on the run. When it produced a form correctly, this means that the combination of that form's bundle assignment, and all applied phonology produced the output actually observed in English. In addition, like humans, the model produces a regular past for novel words most of the time. Nine out of ten individual runs of the model produced regulars more than half the time (ranging from 67% to 100% regular, mean of 86%). On one run, regulars were produced only 36% of the time.

Note that the model contains no explicit mechanism for preferentially assigning novel words to larger groups. The majority of novel words are assigned to the regular bundle not because it is large per se, but because its bundle representation is not very contentful—it is very close to the 'average' of the language. This is because it is large, *and* because there are no regularities within its members.

---

[2] These were their 'core' verbs, which were designed to test hypotheses about islands of reliability. They tested an additional 17 'peripheral' verbs which we do not test.

Operation	No. forms	Example forms
FAITHFUL	35	hurt, split, shed, bet, trust, ...
i → ε	13	lead, feed, read, ...
ɪ → æ	11	swim, shrink, drink, ...
ɪ → ʌ	11	sting, stick, cling, ...
a͡i → o͡ʊ	11	stride, drive, rise, ...
i → ε, ADD –/t/	8	deal, mean, keep, ...
e͡j → ʊ	6	take, mistake, partake, ...
i → o͡ʊ	6	steal, speak, freeze, ...
d → t	5	build, spend, send, ...
a͡i → ɪ	4	slide, hide, bite, ...
a͡i → a͡ʊ	5	grind, wind, find, ...
ε → o͡ʊ, ADD –/d/	4	tell, sell, retell, ...
o͡ʊ → u	5	blow, throw, grow, ...
u → ɪd	4	do, redo, outdo, undo
ε → o͡ʊ	3	hold, uphold, behold
ε˞ → ɔ˞	4	swear, bear, tear, wear

Figure 7: Irregular bundles learned by the model: Their contents and affiliated operational constraints

The model categorized the irregular forms into the same set of bundles on each run. Example bundles with their operational constraints are given in Figure 7.

The regular past was represented differently from run to run. On eight out of ten runs, a single regular bundle was formed. The most common operational constraint affiliated with this bundle was ADD – əd, which occurred on six runs. In this case, the phonological component of the model included a high-weighted *ə constraint so that there was general schwa deletion blocked by a constraint against two adjacent coronal stops. The constraint against voicing disagreement word-finally also had a high weight so that [d] would be devoiced to [t] after a voiceless consonant. Schwa deletion does not obtain in general in English, but the learner was not given enough data to learn that. We predict that if the learner were also given correct outputs like *bracket* realized as [bɹækət], it would choose a different underlying form for the regular past. On one run, the learner chose –/t/ as the underlying form of the regular, and on one run it chose –/d/ as the regular.

On two runs, the learner formed three distinct bundles for the regular: coronal stop-final words affiliated with ADD –əd, voiceless-final words affiliated with ADD –t, and voiced-final words with ADD –d. On these runs, no phonological markedness constraints were weighted high enough to change the output of the operational constraints. In this case, the assignment of a wug verb to a particular past tense category was based on the similarity in the same way that assignment to an irregular category was. The result was phonotactically illegal outputs for some wug words, such as [ʃɪlkd], or [dɹɪtd].

Figure 8 shows how the model behaved on words in different Islands of Reliability. Each bar represents the proportion of productions of regular (or irregular) past tense forms for bases in each type of IOR. Past tense forms can match the IOR of their base (e.g. an irregular past for a base in an irregular IOR), or mismatch the IOR of their base (e.g. regular past for a base in an irregular IOR). Additionally, some bases are not in any IOR's and these are represented with their own bars on the graph. Regular productions are more likely than irregulars, even for bases in irregular IORs. However, bases in both regular and irregular IOR's show a bias to take a past tense corresponding to their IOR. Past tenses which contradict the IOR of their base are least common, and an intermediate amount of regular/irregular pasts is produced for bases in no IOR.

The distinction between matching and mismatching for regulars is not necessarily reliant on the statistics of regulars *per se*. Instead, it is conceivable that the difference emerges because regular IOR forms simply lack viable competition with irregulars. As such, their probability could be enhanced without any sense of a regular IOR in the model itself. However, the non-IOR regular productions suggest that regular IORs may have an effect emerging from the statistical properties of lexical items in English. There will typically only
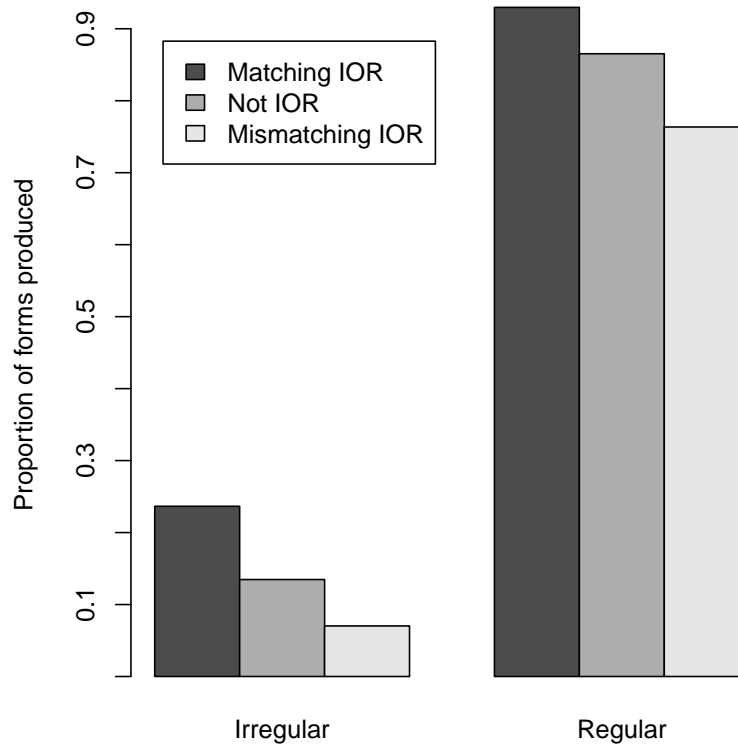
Figure 8: Proportion of regular or irregular forms within a type of Island of Reliability. A matching IOR production is one in which a regular is produced for a lexical item in a regular IOR (*cet. par* for irregulars). In a mismatching production the IOR does not promote the production.

be *one* regular bundle, meaning that any IOR effects have to be created by the mean violations of that bundle in comparison with the mean violations of the irregular bundles. Even with this limitation, regular IORs are seemingly more typical of regulars in the sense that they are closer to the bundle mean. This typicality boosts their probability above the overall boost to regulars for being outside of irregular IORs. Typicality in mean violation is a different notion than IOR and this distinction requires further study.

5   Conclusions

In this paper, we presented a model in which a structured lexicon interacts with morphological operations and phonological constraints. Phonological markedness and faithfulness constraints compete together with operational constraints dictating morphological exponence. These operational constraints are specific to particular bundles of lexical items, and bundle membership is determined during the learning process by bases' relative similarity. Similarity is calculated using markedness constraint violations. A morphological 'rule' is therefore decomposed into two parts: an operation, dictating the change which the base must undergo, and the mean violations of the bundle, approximating a context of application. This separation of the change in a morphological rule from its context of application is similar to the 'phonotactics over sublexicons' approach proposed by Gouskova & Newlin-Łukowicz (this volume).

This model has the advantage that it can learn morphological rules and phonological constraints side by side, and the model's choices about one system can influence the other. For example, when the model settles on a single morphological operation for the regular past, it also learns phonotactic constraints which cause phonological changes to the output of the rule, leading to voicing agreement and schwas separating coronal stop-stop sequences. When it learns three separate morphological operations for the English regular,

it fails to learn these phonotactic constraints since they are redundant with the effects of the operations. We speculate that factors such as the generality of the phonotactic patterns in the input data might condition which combination of morphological operations and phonological constraints the model settles on.

Because the model jointly learns phonological constraints and morphological operations, it is sensitive to morphological subgeneralizations, but those subgeneralizations do not necessarily override productive phonotactics. We demonstrate that the model learns sensible subgeneralizations in the English past tense system, and that like humans, it respects Islands of Reliability for both regulars and irregulars.

The basic structure of this model offers possibilities for effectively modeling various types of phenomena. The German plural system closely parallels the English past system in that it has a default form, and exceptions which pattern together phonologically. However, unlike the English past, the German plural's default is relatively infrequent (e.g. Marcus et al., 1995). In our model, the infrequent default would be affiliated with a bundle whose membership is small but diverse. Because its membership is diverse, novel words that did not conform closely to other bundles would be most similar to the default. Thus we would expect results similar to those reported in this paper, namely that most novel words would be inflected with the default. Only words in islands of irregular islands of reliability would be inflected as irregulars.

In our model, the target of a morphological operation need not be local to its conditioning context, and the conditioning features or segments need not be local to each other. Because of this, we aim to model patterns like the Arabic broken plural (McCarthy & Prince, 1990), in which various non-adjacent features in the base (for example features in both the onset of the first syllable and nucleus of the last syllable) jointly condition the choice of a templatic operation.

In the English past case, several phonological properties work together to condition a single morphological decision. However, it would be possible to extend the model minimally so that a single bundle could be indexed to an array of morphological operations, or even phonological constraints. This would allow for modeling of lexical strata, such as those in Japanese, in which each strata has an array of characteristic phonological properties, and an array of characteristic morphological and phonological behaviors (Ito & Mester, 2002). Another straightforward extension of the model would be to allow non-phonological properties to be incorporated into bundle representations. This would allow properties such as a word's syntactic category to condition the application of a process. Such a model could capture patterns such as English stress, in which a word's syntactic category conditions the outcome (Guion et al., 2003).

References

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

Alderete, John (2001). *Morphologically governed accent in Optimality Theory*. Psychology Press.

Aronoff, Mark & Zheng Xu (2010). A realization optimality-theoretic approach to affix order. *Morphology* 20:2, 381–411.

Baayen, R Harald, Richard Piepenbrock & Rijn van H (1993). The CELEX lexical data base on CD-ROM .

Boersma, Paul (2001). Phonology-semantics interaction in ot, and its acquisition. *Papers in Experimental and Theoretical Linguistics* 6, 24–35.

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

Clahsen, Harald, Ingrid Sonnenstuhl & James P. Blevins (2003). Derivational morphology in the German mental lexicon: a dual mechanism account. Baayen, Harald & R. Schreuder (eds.), *Morphological structure in language processing*, Walter de Gruyter, Berlin, 125–155.

Goldwater, Sharon & Mark Johnson (2003). Learning ot constraint rankings using a maximum entropy model. Spenader, Jennifer, Anders Eriksson & Osten Dahl (eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, 111–120.

Golston, Chris (1996). Direct Optimality Theory: Representation as pure markedness. *Language* 713–748.

Gouskova, Maria & Luiza Newlin-Łukowicz (this volume). Phonological selectional restrictions as sublexical phonotactics. Kingston, John, Claire Moore-Cantwell, Joe Pater & Robert Staubs (eds.), *Proceedings of the 2013 meeting on phonology*.

Guion, Susan G., J.J. Clark, Tetsuo Harada & Ratree P. Wayland (2003). Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech* 46:4, 403–427.

Hayes, Bruce (2004). Phonological acquisition in Optimality Theory: the early stages. Kager, Rene, Joe Pater & Wim Zonneveld (eds.), *Fixing priorities: constraints in phonological acquisition*, Cambridge University Press.

Ito, Junko & Armin Mester (2002). One phonology or many? issues in stratal faithfulness theory. *The Phonological Society of Japan (ed.) Phonological Studies* 5, 121–126.

Jäger, Gerhard (2007). Maximum entropy models and stochastic Optimality Theory.

Marcus, Gary F., Ursula Brinkmann, Harald Clahsen, Richard Wiese & Steven Pinker (1995). German inflection: the exception that proves the rule. *Cognitive Psychology* 29, 189–256.

McCarthy, John J. (2000). Harmonic serialism and parallelism. Hirotani, Masako (ed.), *Proceedings of the North East Linguistics Society 30*, Amherst, MA, 501–524.

McCarthy, John J (2007). *Hidden generalizations: phonological opacity in Optimality Theory*. Equinox.

McCarthy, John J & Alan S Prince (1990). Foot and word in prosodic morphology: The arabic broken plural. *Natural Language & Linguistic Theory* 8:2, 209–283.

Nosofsky, Robert M. (1990). relations between exemplar-similarity and likelihood models of classification. *journal of mathematical psychology* 34, 393–418.

Pinker, Steven (1999). *Words and Rules: the ingredients of language*. William Morrow, New York.

Pinker, Steven & Alan Prince (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28:1, 73–193.

Prasada, Sandeep & Steven Pinker (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes* 8, 1–56.

Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Blackwell, Malden, MA and Oxford, UK.

Ratcliff, John W, David Metzener et al. (1988). Pattern matching: The gestalt approach. *Dr. Dobb's Journal* 13:7, 46–72.

Rumelhart, D. E. & J. L. McClelland (1986). *On Learning the past tenses of English verbs*, MIT press, chap. 18, 216–271.

Staubs, Robert (2011). Operational Exponence: Process morphology in Harmonic serialism. *Challenges of Complex Morphology to Morphological Theory, Linguistic Society of America Summer Institute, Boulder, CO. Handout.* .

Wilson, Colin (2013). A targeted spreading imperative for nasal place assimilation. Fainleib, Yelena, Nicholas LaCara & Yangsook Park (eds.), *Proceedings of the 41st meeting of the North East Linguistics Society (NELS 41)*.

Wolf, Matthew Adam (2008). *Optimal Interleaving: Serial phonology-morphology interaction in a constraint-based model*. Ph.D. thesis, University of Massachusetts Amherst.