Paradigms Learned One Word at a Time and their Structure*

Ehren M. Reilly

The Johns Hopkins University

1. Introduction

Although there is widespread agreement that the morpho-phonology of paradigms involves some sort of process being applied to some sort of base, there has been considerable debate as to what kinds of representations function as morpho-phonological bases, and what processes are involved in generating surface forms from bases. More worrisome than the diversity among the proposals is the extreme disagreement among theorists about the relevant criteria for evaluating the goodness of a theory of lexical representation and morpho-phonology. The purpose of this paper is to pursue clarity on the issue of the appropriate desiderata for a theory, and to develop a rough sketch of a new model for lexical representation, learning and morpho-phonological grammar.

2. Theories of Morpho-phonological Bases

There are generally two dimensions along which theories of bases diverge: abstractness and economy. This yields four extremes, of which three are exemplified by fully articulated theories. Perhaps the simplest view is that bases are concrete, attested surface forms, and that all known surface forms are stored in the lexicon. At the opposite extreme, the lexicon is maximally economized, and this is achieved by contriving abstract, succinct underlying representations. Albright (2002b) upholds the principle of economy but eschews abstractness, proposing that a single member of the paradigm serves as the base for the rest. To my knowledge, there is no articulated proposal for the fourth extreme, in which multiple abstract UR's are stored as bases for a single paradigm.

2.1 Problem with Underlying Representations: Paradigm Uniformity

A major source of traditional linguistic evidence for rich surface representations has been

* I would like to thank Adam Albright, Bill Badecker, Luigi Burzio, Joan Chen-Main, Sara Finley, Kathryn Flack, Gaja Jarosz, Paul Smolensky, and all the visitors to my poster for their insightful comments and productive discourse.

Paradigm Uniformity (PU) phenomena (Burzio, 2005, McCarthy, 2005). Under PU, surface forms in a paradigm appear to be faithful to or based on each other, rather than to their putative shared underlying representation. Hayes (1999) points out that there is often more predictability and greater potential for descriptive generalization to be found between surface forms within a paradigm than between the set of surface form and their UR. These findings suggest that URs are not rich enough to adequately describe the structure of paradigms, because they miss pervasive surface-to-surface relations.

2.2 Problem with Multiple Bases: Unidirectional Diachronic Analogy Gradient

A sensible treatment of PU is to postulate that all surface forms are stored in the lexicon, and that the similarities among surface forms are the result of faithfulness constraints enforcing similarity between them (Burzio, 2005, McCarthy, 2005). However, this assumes that all surface forms in the paradigm have essentially equal status. This seems intuitively unlikely, as certain forms seem more prominent, less marked, more basic, etc. than others. As Albright (2002b *et seq.*) notes, this is not just an intuition, but a fact that is well documented in the historical linguistic record. Analogical diachronic changes overwhelming analogize towards a single cell of the paradigm (Kiparsky 2006). This means that one member of the paradigm has a special privileged status not afforded to the others. Albright identifies this special form as the unique base of the paradigm, and provides a wide range of historical and experimental evidence suggesting that in a particular language, at a particular time, one single surface form must rule all the others.

2.3 Problem with Single Surface Base: Neutrast Configuration

When a phonological contrast is preserved in some members of the paradigm but not others, the contrast-preserving forms are crucially informative. A UR-less lexicon is descriptively inadequate if it does not include these crucially contrastive forms. In Hayes' (1999) discussion of multiple predictability within paradigms, he suggests that "speakers memorize **enough** inflected forms in each paradigm to be able to project the others; specifically, in cases of phonological neutralization they often need to memorize at least one allomorph that escapes the neutralization." But how many is enough? More pointedly, does Albright's parsimonious single base always contain enough information to adequately project the rest of the paradigm?

Perhaps surprisingly, one base is usually enough. Because Albright's Minimal Generalization Algorithm (MGA) is a formal learning algorithm, we can actually assign a quantity to its descriptive adequacy. Even for the most irregularity-laden, morphological complex languages, the grammar learned by the algorithm typically achieves error rates of less than 10%. Mutual predictability is often absolute, meaning that **any** member of the paradigm would be an informationally adequate unique base (Hayes 1999). Where mutual predictability fails, there are often sub-regularities that cluster into "islands of reliability", which a sophisticated enough grammar can exploit (Albright, 2002a). Suppletive morphology can be encoded in a single-base model by highly specific suppletive mappings. However, there are still situations in which a single base is quite obviously the wrong solution.

An example of this failure is found Texistepec Popoluca (TP), a Zoquean language of southern Mexico. Several members of the inflectional paradigms for nouns and verbs in TP include what is at least historically a prefix /j-/.[1] The morpho-phonemic alternations between "palatalized" (*j*-full) and "non-palatalized" (*j*-less) forms include a variety of processes effecting the initial consonants of stems, and a pair of vowel alternations. Of interest to us are the vowel categories /ɨ/, /i/ and /e/. In non-palatalized forms, [ɨ] contrasts with [i], but in palatalized forms this contrast is neutralized to [i] (1a,b). At the same time, in palatalized forms, [i] contrasts with [e], but in non-palatalized forms, this contrast is neutralized to [e] (1c,d). Each type, palatalized and non-palatalized, preserves the contrast that is neutralized in the other—a "neutrast" configuration.[2]

(1) Texistepec Popoluca Paradigms with [ɨ], [i], and [e].

		(UR)	3RD ERGATIVE (palatalized)	3RD ABSOLUTIVE (non-palatalized)
a.	'fall'	/pɨŋ/	[piŋ]	[pɨŋ]
b.	'pinch'	/piŋ/	[piŋ]	[peŋ]
c.	'chicken'	/pij/	[pij]	[pej]
d.	'sway'	/pej/	[pjej]	[pej]

This situation is easily described in a UR-based theory by positing underlying phonemes /ɨ/, /i/ and /e/, but for Albright's single base surface-to-surface approach, these paradigms are highly problematic. If the unique language-wide base is a *j*-full form, then there is no sensible way to encode the [ɨ]~[i] contrast—all incidence of surface [ɨ] would be treated as irregular allomorphy of [i], despite the fact that the [ɨ]~[i]~[e] neutrast is perfectly regular. If a non-palatalized allomorph is chosen as base, a similar problem arises with the [i]~[e] contrast. In order to adequately project both the palatalized and non-palatalized allomorphs from surface representations alone, a TP speaker's lexicon would need to include both a palatalized and a non-palatalized base. So how many bases are "enough" in TP? The answer appears to be exactly two.

When considered altogether, the realities of PU, unidirectional diachronic analogy, and neutrast configurations are not entirely compatible with any existing model of lexical representation and lexical learning. The storage of all surface forms as bases, in addition to possibly being overkill, fails to account for the fact that one base is always

---

[1] (2) shows the full set of Texistepec Popoluca inflectional prefixes from (Reilly, et al., in prep.).

(2)

	Ergative	Absolutive
1st	/n-/	/k+/
2nd	/j-n-/	/k+j-/
3rd	/j-/	Ø

[2] The categories /ɨ/, /i/ and /e/ are all pervasive, appearing, respectively, in the alternating position of 6.5%, 10.0%, and 9.9% of the nouns and verbs in Reilly et al. (in prep.).

far more salient than the rest. Yet a single surface form or underlying representation as base proves to be descriptively inadequate. An adequate model of morphological paradigms must identify a primary base without excluding from the lexicon other bases that may be crucial to the accurate projection of the rest of the paradigm. In this paper, I propose a model that learns and uses paradigms with such a structure. Before articulating this model, I believe it is important to consider what else we should expect from a model of morphological paradigms, aside from explaining the data already discussed above.

3. Additional Desiderata for a Theory

What are the relevant considerations for the design of a model of paradigms? When proposing a model, it is crucial that one be clear about what exactly one is modeling. Ideally, a theory of paradigms and the lexicon ought to be compatible with data from ordinary morpho-phonological description, acquisition data, data from diachronic change, and data from psycholinguistic experimentation. Towards this goal, I have amassed the following desiderata for theory of the lexicon, as a supplement to those already considered above.

3.1 An Explicit Model of Morpho-phonological Acquisition

Especially since the inception of Optimality Theory, phonologists have been aware that a theory of phonological grammar can and should constrain and inform theorization about the acquisition of lexical representations and of phonology. There is now an extensive literature on phonology and morphology learning algorithms, and a general consensus that accurate grammar description and learnability must go hand-in-hand.

3.2 Compatibility with Gradual and Lifelong Lexical Learning

Although there is a critical period for the acquisition of core grammar, the acquisition of new words, including sets of words that exhibit morpho-phonological alternations, proceeds well into adulthood. There is no end-point for lexical learning. Nor is there a discernable discrete beginning. Children begin using the grammar and lexicon long before they are fully learned. Many of today's best phonology learning algorithms have a beginning and an end point, and neither the grammar nor the lexical items are available for use during the learning process. For example, Tesar (this volume) employs lexical representations with temporarily "unset" features during the learning process, which he indicates with the symbol '?'. According to Tesar, "This symbol only has status with respect to the learner, not with any adult phonology itself." Tesar probably does not actually believe that lexical representations are unavailable for use by the speaker/learner until enough members of the paradigm have been encountered to fully determine the UR. More likely, batch learning just makes it easier to train and test the algorithm. A more sophisticated version of Tesar's model would need to say something about how a young child might pronounce an underlying /?/. Although this detail may rightly be overlooked for the sake of simplicity, we should generally avoid models that would be incompatible with an upgrade to the continuous time course of acquisition.[3]

3.3 Explicitness about the Deletion of Unnecessary Representations

The lexical memory capacity of the human mind is, of course, finite. The generative capacity of inflectional morphology is immense, and, especially in languages with very rich inflection, it is certain that the range of felicitous words exceeds our memory capacity. However, the desire among generative linguists for lexical storage to be as small as logically possible is driven largely by the quest for an elegant theory, not by empirical data. Several authors on this topic have questioned the relevance of this factor, and have shown psycholinguistic evidence that, for better or for worse, human speakers actually do memorize certain aspects of predictable surface forms (Baayen et al., 1997, Bybee, 1995). Others have argued for the necessity of mass storage of surface forms on theoretical grounds (e.g., Burzio, 2005). Economy is, in some sense, a relevant consideration for theory building, but there is no empirical basis for its being paramount.

A process that is implicit in economization, which has been largely overlooked by linguists is the process of forgetting, decay, or deletion whereby lexical representations are purged from memory. Under any theory, the learning process for phonological or morpho-phonological alternations involves accurately identifying surface representations and making some sort of inference based on them. For example, learning the correct underlying representation for a morphological paradigm often requires simultaneously comparing several members of the paradigm. Since it is unlikely that real children are presented with the crucially informative surface forms all at the same, they must store these surface forms in memory in order to make complex inferences later on. How long are these surface representations stored? Are they stored in a special learning buffer, or are they stored in the same lexicon as the eventual underlying representations? If, as stipulated by the underlying representation hypothesis, all surface forms are eventually purged from memory, when does the purging happen? How do learners know when they have extracted all the information that they need to know from the surface forms in order to be able to abandon them from memory? Certainly, there cannot be one single moment labeled "morpho-phonological maturation" at which the rich inventory of surface representations is analyzed and transformed into an economized inventory of bases. The learning and decay process must be dynamic, both because lexical learning continues well into adulthood, and also because the availability of the crucial information differs greatly between words depending on their frequency and/or age of acquisition. Also, if surface lexical representations are deleted once they have been obviated by some abstraction or generalization, is this deletion process the same as that of forgetting a rare word due to lack of use? A viable theory of the lexicon must address these issues and

[3] I will not consider in this paper the issue of whether the child's phonological representation is a featurally specified as the adult's. There is now some evidence that the amount of featural information that children can store in a lexical representation increases gradually with age, and that very young children may have impoverished lexical representations with room for only one place feature (Fikkert and Levelt 2006). This is an important consideration, and I believe that the general approach I develop in this paper is consistent with this possibility. I will, however, assume for the sake of simplicity that there are no qualitative differences between the lexical representations of adults and of children.

make explicit, testable assertions about the economization, forgetting, or deletion process.

3.4 Matching and Predicting Speakers' Intuitions about New Words

To claim that a speaker's morpho-phonological grammar and lexicon are organized in a particular way makes an implicit prediction about how that speaker's grammar will treat a new and unfamiliar lexical item. Since Berko (1958), a common psycholinguistic method for testing morphological knowledge has been the so-called "*wug*-test," in which subjects are presented with one member of a novel paradigm and asked to project from it another member of the paradigm. Particularly for the case of English tense morphology there is an extensive experimental literature that develops theories of morphological grammar based on nonce-probe experiment data (Berko, 1958, Bybee and Moder, 1983, Prasada and Pinker, 1993). Particularly with the advent of algorithmic and connectionist computational models, the ability of a model to not only produce the attested forms of the language but generalize to nonce forms in the same way as humans has become an important test of a formal model's psychological plausibility (Albright and Hayes, 2003, Bybee and Moder, 1983, Prasada and Pinker, 1993).

4. Proposed Model

In this section I will develop a model of lexical learning and lexical representation that I believe can account for all of the data in Section 2. As for the desiderata in Section 3, most existing models are not comprehensive and explicit enough with respect to these criteria to even make evaluation possible. The only exception to this generalization is Albright's MGA. For this reason, I will take the MGA as a starting point, and aim to improve upon its failures without sacrificing its successes.

4.1 Sketch of the Original MGA

The Minimal Generalization Algorithm is described in detail in Albright (2002b) and in Albright & Hayes (2002). Since the only data in this paper are from TP, the brief sketch of the algorithm given here will use TP items as examples.

The MGA begins by identifying surface words determined to be morphologically related. For the algorithm to work, not all surface forms within a paradigm need to be attested, but all attested forms are compared to create a set of every pair of morphologically related forms.

Step 1: Identify Morphologically Related Ordered Pairs

1ERG → 2ERG		3ERG → 3ABS		3ABS → 3ERG		
m͡beŋ$_{\text{1ERG}}$	m͡biŋ$_{\text{2ERG}}$	piŋ$_{\text{1ERG}}$	peŋ$_{\text{2ERG}}$	peŋ$_{\text{2ERG}}$	piŋ$_{\text{1ERG}}$	…
w̃ẽn$_{\text{1ERG}}$	w̃j̃ẽn$_{\text{2ERG}}$	wjen$_{\text{1ERG}}$	wen$_{\text{2ERG}}$	wen$_{\text{2ERG}}$	wjen$_{\text{1ERG}}$	…
naj$_{\text{1ERG}}$	ɲaj$_{\text{2ERG}}$	dʲaj$_{\text{3ERG}}$	daj$_{\text{3ABS}}$	daj$_{\text{3ABS}}$	dʲaj$_{\text{3ERG}}$	…

Then, the MGA compares these pairs of strings in order to identify a mutated region, labeled "change" below, and the unchanged "environment" regions surrounding it. A single directional rule is then posited for each of these changes.

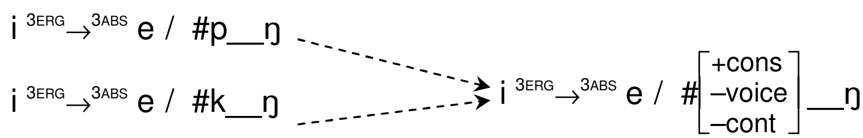
Step 2: Compare Strings and Identify Changes in Environments

Step 3: Posit a Specific Rule for Each Change

n	1ERG→3ABS	d	/ #__aj#	dʲ	3ERG→3ABS	d	/ #__aj#
m͡b	1ERG→3ABS	p	/ #__eŋ#	i	3ERG→3ABS	e	/ #p__ŋ#
w̃ẽ	1ERG→3ABS	we	/ #__n#	wj	3ERG→3ABS	w	/ # __en#

Then, generalizations are made between rules that perform different changes in the same environment. These generalizations are "minimal" in the sense that the generalized changes are as specific as possible, given the set of strings changes being combined. This process repeats iteratively, until no additional generalizations can be made.

Step 4: Iteratively Generalize Rules



A fifth step involves subtracting from these morpho-phonological relations any relations that can be attributed to purely phonotactic restrictions (see Albright 2002b).

The final step is to calculate the reliability of each of these rules. When the algorithm is used to generate forms, there will be many different rules which may apply in the given context. These rules will differ in their specificity and in their overall accuracy. The selection of the best rule will depend on both of these factors, via a reliability statistic that is computed for each rule. In the MGA, reliability calculations based on more data get a higher confidence rating. The penalty for smaller generalizations is an adjustable parameter in the model (see Albright, 2002b, Albright and Hayes, 2002).

Step 6: Compute Reliability for All Rules

Accuracy = **# of hits** (known changes correctly predicted by rule) / **scope** (environments in lexicon where rule could apply)

Reliability = A statistically adjusted function of Accuracy

Once these statistics have been computed, the learner has a set of surface forms, and a large set of reliability-weighted directional rules for deriving new surface forms. For Albright, there is an additional stage of the model, in which, based on the overall reliabilities of the mappings from each member of the paradigm, across the entire lexicon, a single paradigm member is chosen as the unique base. At this point, all non-base surface representations and the rules to derive other surface forms from them are deleted. What remains is a set of all forms belonging to a particular paradigm slot, and the set of all the rules that map from this slot to another. Non-base surface forms in a paradigm are projected from the base of the paradigm via application of the most reliable available rule.

4.2 Advantages of the Original MGA

Research on this algorithm has yielded a variety of compelling results. First, the storage of only a single surface form for each paradigm is as economical as a surface-to-surface morphology system can be. And even with this minimal storage, the descriptive accuracy achieved by the grammars that MGA is typically over 90%. But in addition to being economical, and fairly descriptively accurate, the MGA demonstrates an unprecedented predictive power in the domains of predicting and modeling analogical diachronic change, and correctly modeling the responses of participants in *wug*-test experiments.

The choice of a single-base representational system was motivated by the finding that analogical changes only ever analogize to one member of the paradigm, across the language as a whole. Which member this is may vary between languages, and may change with time as a language changes, but the trend within a language is unidirectional at any given time (Albright, 2002b, Kiparsky, 2006). The MGA is, among other things, an algorithm for identifying this privileged slot in the paradigm. An example taken up by Albright (2002b, 2005) is the leveling of a paradigmatic alternation in Latin by analogizing the nominative “citation” form to a more phonologically complex oblique case. Albright demonstrated that at this particular time in Latin, the nominative was less reliable than the oblique cases in terms of intra-paradigmatic predictability, according to the MGA’s metrics. The MGA’s modeling of typologically unusual paradigmatic effects in Yiddish and Lakhota is equally compelling (Albright 2002b).

The calculation of a rich set of surface-to-surface string mappings, which vary in their specificity and reliability, allows the grammar to encode both productive and unproductive sub-regularities. There is considerable experimental evidence from *wug*-test experiments that speakers have access to paradigmatic sub-regularities (Bybee and Moder, 1983), so an adequate model must capture this information. The MGA has been shown to do this for Italian (Albright, 2002a) and English (Albright and Hayes, 2003).

4.3 Failures of the Original MGA and Proposed Improvements

Neutrasts like the TP data in Section 2.3 are a serious obstacle for the MGA. But, even in the relatively successful published implementations of this model, the descriptive adequacy of the final product is questionable: 90% accuracy is good, but it leaves some room for improvement. This should not be viewed as a fatal flaw, of course, since this is actually a very nice result by computational learning algorithm standards. Furthermore, since no one has succeeded in implementing one of the competing architectures to do this task, we can hardly criticize the MGA for finishing in first place with an imperfect score. Instead, let us ask exactly how and why the MGA falls short of the mark, and what kinds of modifications would improve the overall performance without losing any of the positive attributes.

What is immediately troubling is that the MGA deletes most of the information that is collects. The process of unique base selection involves a rather inelegant one-time purging of all rules and representations pertaining to the less reliable paradigm slots. After going to the trouble of computing rules and reliabilities for all directional pairwise surface-to surface relations, the MGA selects a single base, and abandons all other bases, along with all of the rules that would project other surface forms from them.

In general, batch learning not an ecologically valid model for language acquisition, and purging unnecessary representations at the "end" of learning is clearly an implausible approach to forgetting and decay. More likely, each time we hear a new word, we iterate the processes of storing the relevant lexical representation, computing morpho-phonological relations, generalizing these relations, and integrating this information into out our active grammar hypothesis. A superior model of acquisition would not draw a temporal distinction between learning words and using them. As for decay, it is unclear whether this is an active process that is executed every time we invoke our lexicon, or a passive process that happens gradually all the time. Either way, it, too, must be spread across most of the lifetime.

Descriptively, the only problem with unique base selection is the information that is lost. In TP, the selection of either a palatalized or non-palatalized form as the unique base neutralizes a contrast that was clearly encoded in the grammar prior to base selection. What would compel a real-life learner to adopt only the *j*-form as base, to the exclusion of the crucially informative Ø-form, especially when the learner already knew—and presumably used—this Ø-form prior to the completion of learning? Even if there were an active pressure to forget as many surface forms as possible, a TP learner must preserve ate least two different surface forms of the [ɨ]~[i]~[e] words just to stay afloat.

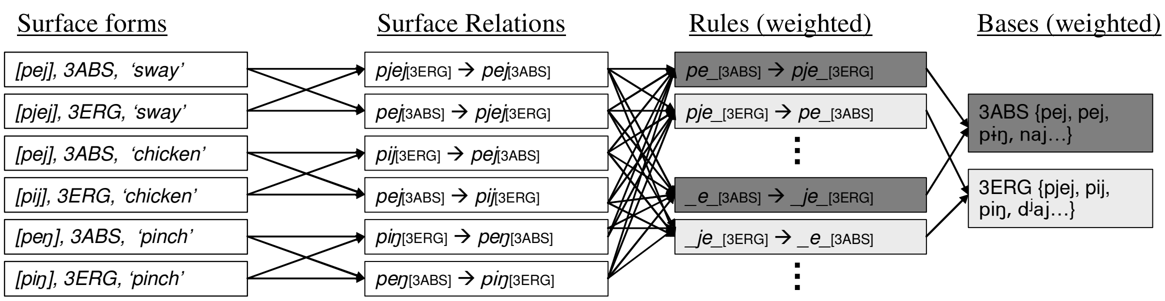
4.4. Learning Multiples Bases

The original MGA learner applies each learning step as a function over the complete lexicon, all at once. I propose an algorithm which iterates essentially the same functions each time a new surface form is encountered. Exposure to new forms introduces new surface relations. That is, each time a new word is learned, it is compared with the existing lexicon. From these comparison, the learner creates a set of rules, weighted for their reliabilities. As new forms are added to the lexicon, this changes the hits and scope of the existing rules, and this change is then reflected in their updated reliability score. New lexical material will often necessitate the introduction of new rules, which can proceed iteratively just as in the original MGA.

As for base selection, the original MGA chooses only one, and does this across the entire lexicon all at once. In the version I propose, each potential base in the lexicon is assigned a strength, according to the metrics in Albright (2002b). However, weaker bases are not necessarily eliminated, they are simply assigned less prominence.

The figure in (8) illustrates how the proposed model might treat the 3rd absolutive and 3rd ergative forms of three words. On the left is the set of attested surface forms. Computations over these surface forms yield a set of rules, weighted for reliability, and this set is updated upon the acquisition of each new form. The relative prominence of each base in the paradigm is a function of the reliability of the rules and the lexicon.[4] In this figure, weaker and less reliable representations are more darkly shaded.

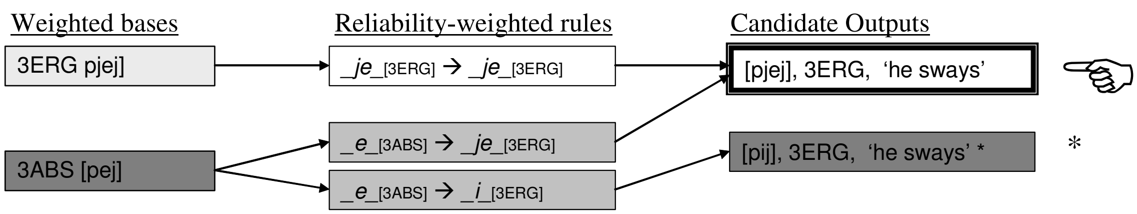
(8) Word-by-Word Learning in a MGA with Gradient Non-unique Basehood



4.5 Projecting Surface Forms from Multiple Bases

Under this new model, the lexicon consists of a rich set of surface representations, weighted according to their prominence as bases. In the example in (8), the 3rd person ergative is the most prominent "strong" base, while the 3rd person absolutive is the less prominent "weak" base. The projection of surface forms in this model involves invoking all stored bases that belong to a particular paradigm. As in the original MGA, morpho-phonological rules are applied to these bases, and these rules project outputs. In this new model, outputs are weighted according to the product of the base strength and the rule reliability. That is, stronger bases and more reliable rules both contribute to the weight of the output. Because weak bases have less activation to send, they are rarely able to influence the choice of output. As illustrated in (9), the uninformative weak base does not interfere with the projection of surface forms.

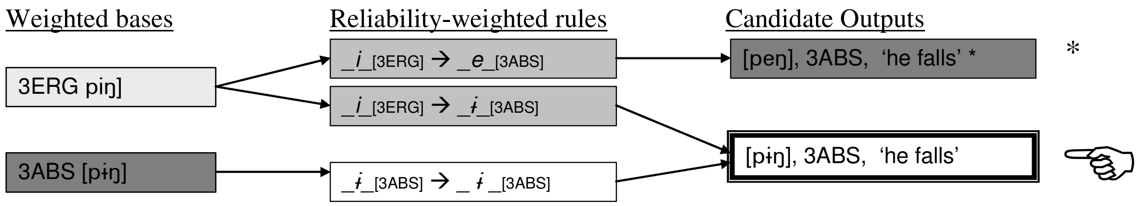
(9) Projecting Surface Forms When the Strong Base is Reliable



[4] The exact scoring system that should be used to rank bases is a matter of uncertainty, and Albright (2002b) considers a number of different possibilities, most of which lead to the same results.

However, since the strong base is palatalized, correctly projecting a surface [ɨ]~[i] contrast in non-palatalized environments will crucially depend on the contribution of the weak base. When the strong base is unreliable, the total activation it sends to its candidate outputs is lower, and, if the weak base is reliable, it will send a significant enough quantity of activation to its favored output to tip the scales in the right direction, as in (10).

(10) Projecting Surface Forms When the Strong Base is Unreliable



4.6. Attributes of the New Model

At least for these data, the new model achieves superior descriptive accuracy, although a formal implementation of this model has yet to be rigorously tested. Like the original MGA, this model identifies the same single most salient base, which universally serves as the basis for diachronic analogy. No diachronic predictive power is lost in the new architecture. What is gained, however, is increased ecological validity, especially in the domain of economization and forgetting. Under this model, we can postulate that some lexical representations fade away due to lack of use, and we can formally define lack of use in terms of the extent to which representations actively contribute to the projection of other members of the paradigm. The preferred victims for decay are representations that never contribute any crucial information to the paradigm. Depending on how rapid a decay function one deems appropriate, surface forms that are especially predictable or uninformative will eventually fade from memory, leaving only those lexical representations that are necessary to get the job done. Under this approach, economy of representation is not a stipulation, but an adjustable parameter of the model.

5. Conlusions

This paper has motivated and sketched a new approach to morpho-phonological representation and learning, one based on the Minimal Generalization Algorithm. One crucial advantage of this model over other existing models is its ability to adequately describe both paradigm uniformity effects and neutrast situations, while also accurately predicting the direction of analogical changes. An equally important advantage is the explicitness with which this model addresses the realities of gradual lifelong lexical learning and gradual lifelong lexical forgetting. The acquisition of new words and paradigms undeniably proceeds word-by-word over an extended period of time, which necessitates the long term storage of the surface representations that are the basis of paradigm learning. The model proposed here is obviously in its infancy, but it is perhaps the only model in the literature that has attempted to address this broad range of data and ecological constraints. Regardless of the eventual viability of this model, I eagerly hope that the issues I have tried to address will gain increased attention among morpho-phonologists and learning theorists.

References

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

Albright, Adam. 2002b. *The Identification of Bases in Morphological Paradigms*, PhD Dissertation, UCLA.

Albright, Adam, and Hayes, Bruce. 2002. Modeling English past tense intuitions with minimal generalization. Paper presented at *Sixth Meeting of the ACL Special Interest Group in Computational Phonology*, Philadelphia.

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

Baayen, Haarald, Dijkstra, T., and Schreuder, R. 1997. Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language* 37:94-117.

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

Burzio, Luigi. 2005. Sources of paradigm uniformity. In *Paradigms in Phonological Theory*, eds. Laura J. Downing, T. A. Hall and Renate Raffelsiefen. Oxford: University Press.

Bybee, Joan, and Moder, Carol Lynn. 1983. Morphological classes as natural categories. *Language* 59:251-270.

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

Fikkert, Paula, and Levelt, Claartje. 2006. How does place pall into place? The lexicon and emergent constraints in children's developing phonological grammar. Ms. Radboud University of Nijmegen.

Hayes, Bruce. 1999. Phonological Restructuring in Yidiɲ and its Theoretical Consequences. Ms. UCLA.

Kiparsky, Paul. 2006. *Paradigms and Opacity*. Stanford: CSLI Publications.

McCarthy, John J. 2005. Optimal Paradigms. In *Paradigms in Phonological Theory*, eds. Laura J. Downing, T. A. Hall and Renate Raffelsiefen. Oxford: University Press.

Prasada, Sandeep, and Pinker, Steven. 1993. Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes* 8:1-56.

Reilly, Ehren, Kaufman, Terrence, and Bereznak, K. (In preparation). Texistepec Popoluca Lexical Database: Johns Hopkins University.

Tesar, Bruce. (This volume). Learning from paradigmatic information. ROA-795.

Department of Cognitive Science
Johns Hopkins University
3400 N. Charles St.
Baltimore, MD 21218

reilly@cogsci.jhu.edu