AUTHOR QUERY FORM

ELSEVIER	**Journal:** LINGUA **Article Number:** 2059	**Please e-mail or fax your responses and any corrections to:** **E-mail:** corrections.eseo@elsevier.thomsondigital.com **Fax:** +353 6170 9272

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the 'Q' link to go to the location in the proof.

Location in article	**Query / Remark: click on the Q link to go** **Please insert your reply or correction at the corresponding line in the proof**
Q1	Please confirm that given name and surname have been identified correctly. Yes, name and surname are correct. Please check this box or indicate your approval if you have no corrections to make to the PDF file ☐

Thank you for your assistance.

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Lingua

Lingua xxx (2013) xxx–xxx

www.elsevier.com/locate/lingua

Highlights

Lingua xxx (2013) xxx

Linking usage and grammar: Generative phonology, exemplar theory, and variable rules

Gregory R. Guy

Department of Linguistics, NYU, 10 Washington Place, New York, NY 10003, United States

- Abstract phonology uses abstract operations to explain productivity and regularity.
- Usage-based phonology uses memory to explain quantitative detail, lexical diversity.
- Separately, abstract and usage based models each fail to account for some phenomena.
- Variable rule models meld abstract operations with probabilistic quantification.
- The VR approach is capable of modeling the widest range of observed phenomena.

Available online at www.sciencedirect.com
SciVerse ScienceDirect
Lingua xxx (2013) xxx–xxx
ELSEVIER
Lingua
www.elsevier.com/locate/lingua

Linking usage and grammar: Generative phonology, exemplar theory, and variable rules

Gregory R. Guy Q1

Department of Linguistics, NYU, 10 Washington Place, New York, NY 10003, United States

Received 13 September 2011; received in revised form 26 June 2012; accepted 1 July 2012

Abstract

Rule- and usage-based models in phonology are difficult to reconcile: 'rule'-based approaches (including generative and optimality models) rely on abstraction and seek to account for regularity and generality. Usage-based models, like exemplar theory, rely on concrete representations, eschewing abstraction; they typically seek to account for lexically differentiated phonological phenomena, including variability, gradience and probabilistic properties. An alternative that incorporates both generative productivity and quantitative precision is the ''variable rule'' (VR) model of sociolinguistic variation. VR preserves advantages of rule-based models, including abstraction and the capacity to represent categorical processes. But VR resolves many limitations of these formalisms using probabilistic quantification: any phonological process or constraint may be associated with a probability, which permits the treatment of variation and gradience. This paper cites evidence from variation in speech style, child language, and reanalysis across the life span showing that speakers have both discrete abstract and nondiscrete, variable elements of phonology. Variable processes provide a nondeterministic but recoverable link between these different representations.

Keywords: Usage-based phonology; Exemplar theory; Generative phonology; Variable rules; Abstraction; Probabilistic grammar; Productivity; Lexical exceptions; Quantification

1. Introduction

The contemporary debate which is the focus of this issue of *Lingua* opposes 'rule-based' and 'usage-based' approaches to variation and change. The particular instantiations of these approaches that are most commonly referenced in this regard statement are contemporary descendants of generative phonology on the 'rule based' side (cf. Chomsky and Halle, 1968 and its many successors), and Exemplar Theory on the 'usage based' side (cf. Bybee, 2001, 2002; Pierrehumbert, 2006, and many others). However, it is worth noting that this is an old debate in linguistics, particularly in phonology, dating back at least a century and a half. From the Neogrammarians through the structuralists and generativists to Optimality Theory, the mainstream approaches to phonology have been 'rule-based' – this is the family of models that uses some version of the phonemic principle, postulates abstract mental representations of words, spelled out as strings of phoneme-sized segments that distil the phonological essence out of the phonetic details of an utterance, and postulates a phonological grammar in which there occur operations – whether rules or candidate selection by constraint rankings in OT – that capture most generalizable sound patterns of a language.

E-mail address: gregory.guy@nyu.edu.

0024-3841/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.lingua.2012.07.007

Please cite this article in press as: Guy, G.R., Linking usage and grammar: Generative phonology, exemplar theory, and variable rules. Lingua (2013), http://dx.doi.org/10.1016/j.lingua.2012.07.007

Opposed to this, throughout the same span of time, there have been views that we could call 'usage-based', which take a contrary position on some or all of these points. This family of models runs from the anti-Neogrammarian position of the 19th century, which took issue with the regularity of sound change and was typified by the slogan 'each word has its own history' (cf. Jaberg, 1908:6), through lexical diffusion and exemplar theory. In these models, words are represented holistically rather than as strings of segments, typically with rich phonetic detail; phonemic identity is secondary or derivative at best; and the operations of the phonology do not apply across the board to all lexical items, if indeed there are abstract phonological operations at all.

So the issues currently under debate have deep roots; hence we can profitably ask how the history of the debate may illuminate present issues, and for that matter, why this matter has not been resolved before? In my view, the useful insights from this history come from looking at what each line of approach has been good at. The Neogrammarian-descended mainstream argues for 'regular sound change' in the diachronic dimension – ''*die Ausnahmlosigheit der Lautgesetze*'', lit. 'the exceptionlessness of the sound laws'. The synchronic counterpart of this is regular, across the board operation of phonological processes. In the historical record, Neogrammarian regularity is well instantiated: many, probably most changes do not leave historical residues of unshifted segments in exceptional words. Consider for example the proto-Germanic sound change known as Grimm's Law, in which the Proto-Indo-European voiced aspirate stops became plain voiced stops. This did not leave ANY voiced aspirate outcomes in any lexical item in any descendant Germanic language. When these phonological units changed to something else, they were well and truly and completely changed. Similarly, when English lost its voiced velar fricative some seven centuries ago, it lost it completely, leaving not a single lexical item that still sports this once-frequent phoneme, despite the fact that it is still graphemically represented by <gh> in standard English spelling, in bizarre orthographic homage to a sound not uttered by most English speakers lo these 700 years!

The 'rule-based' model developed by the Neogrammarians and still prevalent in phonological theory explains these categorical outcomes by its architecture: the rules and constraints operate on segments and phonological structures, not specific words, and hence all instances of a segment (in the relevant phonological contexts) undergo a sound change or a synchronic operation, regardless of the specific lexical item they occur in. There are further consequences to this architecture. Synchronically, it predicts productivity: speakers know how to pronounce neologisms, loan words, and novel native lexical items for which they have no prior exemplars. It also predicts that speakers can perform abstract phonological operations on entire classes of sounds across the lexicon. But what this model gains in explanatory adequacy, it loses in capacity to account for certain kinds of phonological facts, especially those that involve individual lexical items. Since the mainstream eschews any place for the lexical items in general phonology, it cannot account for lexical frequency effects, lexical diffusion in the course of language change, etc. In fact, the mainstream model predicts that such things should NOT happen.

Remedying these defects of mainstream theory (or MT) is precisely what ET sets out to do. Arguments for exemplar theory famously cite, and seek to account for, the following kinds of phonological patterns:

- Variation: As phonetic and sociolinguistic studies amply demonstrate, phonological units show extensive stochastic variation among assorted realizations in natural speech, well beyond the simple allophonic conditioning envisioned by MT.
- Gradience: Vowel sounds notoriously vary over an articulatory region, and change gradually, passing through a continuum of intermediate points. Exemplarists argue that consonants also show gradated articulations, ranging in the case of lenition from full obstruent realizations to complete absence. MT handles this poorly, with fairly clunky gradual phonetic implementation rules, or not at all, if phonology is construed as consisting exclusively of categorical operations on abstract units.
- Lexical diffusion in sound change: Exemplarists argue that not all sound changes are regular, in the sense of lexically exceptionless; rather, some words may lead while others lag, or do not change at all. As a historical example, note the modern developments of Middle English long /o/. Its regular Great Vowel Shift outcome is /u/, as in *food, mood, root, fool, goose*, but in some words it gets laxed to /U/ (*good, book, soot*), and in other words merged with schwa (*flood*). Some hints of phonetic conditioning are found in these outcomes, but no fully regular statement of context is possible, and there is significant cross-dialect variability.
- Lexical frequency effects: High frequency words are often found to behave differently from lower frequency words. For example, Guy, Hay and Walker's study (2008) of coronal stop deletion in the ONZE corpus of early New Zealand English found that the log of lexical frequency positively correlated with deletion at the $p < .0003$ level: i.e., the more common the word, the more likely it is to appear deleted.
- Common contexts: Words that most often occur in contexts favorable to a particular process undergo that process significantly more often than words that most commonly occur in unfavorable contexts. Thus Guy, Hay and Walker also found that the most common following phonetic context of a word is a significant predictor of coronal stop deletion (with $p < .0002$), independent of whatever the actual following environment is. In MT, the postlexical context of a given utterance of a word should have no influence on how that word is articulated on some other occasion.

Please cite this article in press as: Guy, G.R., Linking usage and grammar: Generative phonology, exemplar theory, and variable rules. Lingua (2013), http://dx.doi.org/10.1016/j.lingua.2012.07.007

Finally, there is one further difference between mainstream phonology and ET, of particular interest to the study of variation and change:

- Probabilistic effects: Studies of natural language systematically demonstrate what Weinreich et al. (1968) have termed 'orderly heterogeneity': quantitative regularities in the frequencies of occurrence of all kinds of linguistic variables, alternations, etc. Essentially, wherever there is an optional element in a grammar – a choice among two or more alternatives – the options are not distributed randomly in natural language usage; rather, they show probabilistic distributions, influenced by linguistic and social contexts. MT has generally ignored such facts, sometimes as a matter of principle (for example, by declaring them to be features of *parole* or performance, for which the theory assumes no responsibility.) Hence MT approaches are typically formulated in categorical terms, with no quantitative capacity. ET, however, proudly asserts its probabilistic capacities (see, for example, Bod, Hay, and Jannedy's (2002) book, *Probabilistic linguistics*), and makes accounting for quantitative patterns a central focus of the model. The extension of exemplarist models to incorporate social indexicality of variables is argued for by, inter alia, Foulkes and Docherty (2006).

But as most readers are no doubt aware, there also exist several probabilistic frameworks derived from MT models that preserve many MT elements. One based on Optimality Theory, known as 'stochastic OT' (Boersma and Hayes, 2001), provides a quantified model that can account for probabilistic patterns of variation. Another, which will be my principle focus here, is a richly probabilistic 'rule-based' alternative to MT, namely the 'variable rule' framework (henceforth VR; Labov, 1969; Cedergren and Sankoff, 1973; Sankoff, 1978; Guy, 1997; Paolillo, 2002). This model grew out of generative phonology and shares many characteristics with MT, including abstract underlying representations and phonological operations on segments. But it shares with ET an interest in accounting for variation, change, and probabilistic patterning, and a quantitative strategy for doing so. Thus VR provides a well-established method for handling at least some of the facts that ET is inspired by. I will argue that the VR model actually provides a superior account of the data, resolving some of the conflicts between MT and ET.

2. Elements of exemplar theory

Exemplar Theory has been designed to account for the kinds of facts cited above that MT does not handle well. How does it achieve these ends? The principle mechanism that differentiates ET from other theories is a greatly enriched representation in memory for lexical items (and possibly also for other items such as morphemes and phonemes, depending on which version of the theory one is considering). Speakers are postulated to retain in memory the exemplars of words that they have heard uttered – in principle, all such exemplars of all words – in high-fidelity phonetic detail, limited only by the resolution of the perceptual apparatus. This cloud of remembered exemplars provides the speaker with information about the pronunciation of individual lexical items, the details of phonetic realization, the patterns of variation evident in the community, and the quantitative distribution of these facts (Pierrehumbert, 2001; Bybee, 2001). This is the crucial difference between ET and MT, and it is the foundation on which the other claimed explanatory abilities of ET rest.

Now there are obvious objections to this basic postulate – it assumes against the evidence that all speakers possess a hyper-eidetic memory for speech (although allowances are typically made for normal memory decay) – but I will pass over this issue and focus on problems that arise in what Exemplar Theory has built on this foundation. Feasting on these greatly enriched underlying memories, Exemplar theorists often conclude that they supply sufficient nutrition to fulfill all phonological needs, so that ET can dispense entirely with such mainstream phonological dishes as abstract operations and representations. Although this sort of theoretical move – going from all of something to none of it – is common in linguistic debates, it runs the usual perils of extremism. In linguistics, as in nutrition, one should perhaps strive for a balanced diet; so before adopting an all memory – no process model, we should examine it more carefully.

Let's consider briefly how ET achieves its ends. The model has three basic elements that do most of its explanatory work. First, the remembered exemplars are stored whole, without being reduced to abstract representations, and without subtracting any allophonic effects of context. They may be 'tagged' in memory as tokens of specific words, and of the speech of particular speakers in particular social contexts, etc., but they are not represented in memory as analyzed sequences of phonemes or syllables. This is a principle of perception that we can call 'Do not decode'.

> Perception Principle: *Don't Decode!*
> Phonetic impressions of heard words are stored without phonological analysis.

Second, certain frequency effects, gradience, and much of lexical differentiation and diffusion are accounted for by what Bybee calls ''automation'' of articulatory routines:

> repeated neuromotor patterns become more efficient as they are practiced; transitions are smoothed by the anticipatory overlap of gestures, and extreme gestures decrease in magnitude or are omitted. (2002:268)

Please cite this article in press as: Guy, G.R., Linking usage and grammar: Generative phonology, exemplar theory, and variable rules. Lingua (2013), http://dx.doi.org/10.1016/j.lingua.2012.07.007

Consequently, sound changes and phonological variants involving lenition will advance gradiently with repetition, such that more frequent words will lead. This is a principle of production, which we can call 'Practice makes (im)perfect'.

> Production Principle: *Practice makes (Im)perfect*
> Frequent repetition of utterances favors gestural overlap and lenition.

Finally, Exemplar theorists such as Bybee and Pierrehumbert argue explicitly that changes progress because of a feedback effect: when some factor (social or linguistic) biases production in a certain direction, exemplars altered in that direction are constantly added into the remembered corpus, which steadily changes the average targets for production, which then suffer further alteration during articulation, and so on, producing a gradual phonetic drift in the direction of bias. This principle, involving the recursive relation between perception and production, we can call 'the feedback loop'.

> Recursion principle: *The Feedback Loop*
> Articulatory reduction produces perceived exemplars that may serve as targets for later articulations which can then undergo further reduction.

Together, these principles achieve most of the results of exemplar theory:

- Variability is present in the input, generated by social diversity plus articulatory variability and preserved in the mental record by Do not Decode.
- Gradience arises from the gradual operation of Practice effects, and is preserved in the mental record by Do not Decode.
- Lexical frequency effects arise because Practice is cumulative and gradient: more frequent words incur more automation, lenition, gestural overlap, etc.
- Change is driven by Practice effects cycled through the Feedback Loop. Lexical diffusion occurs naturally because words are differentiated in usage by frequency of occurrence, and by social and linguistic contexts of occurrence, all preserved by Do not Decode. (NB: Bybee, 2010 and others cite another change mechanism, 'conservation' or 'entrenchment', in which items with high token but low type frequency resist the expansion of generalizing patterns. We do not address this here because it principally addresses morphological or paradigmatic operations rather than phonology.)
- Common context effects arise from Undecoded exemplars biasing the production targets for subsequent productions.
- Probabilistic properties derive from the frequency distributions in the exemplar cloud, as Pierrehumbert (2001, 2002) has spelled out in mathematical models. (Briefly, target selection for production draws stochastically from the exemplar cloud, thus probabilistically replicating the distributions of perceived tokens.)

Thus, ET achieves its ends with rich representations, a set of concrete, primarily physiological processes (of co-articulation, gestural overlap, etc.) that generate diverse realizations, and a set of assumptions about perception, production, and their interaction. These elements give an explanation of all the 'new facts' that ET has focused on. But what about the 'old facts' that MT was designed for? Regular sound change is seen as a long run outcome of the ET mechanisms: since all words containing a given segment in a given allophonic context are subject to comparable articulatory biases, they will all experience Practice and Feedback effects driving toward the same outcome; with the passage of time, sufficient frequencies will be achieved and eventually the outcome of the process may become encoded into every exemplar. Of course, the same model also generates lexical diffusion and irregularity, so the theory does not clearly predict which changes will lead to regularity, and which to lexical diffusion.

Phonological processes that are not easily accounted for as articulatory Practice effects are also a challenge to ET. Lenitions, assimilations, etc. do abound in the languages of the world, but there are also contradictory processes that are not easy to explain as gestural weakenings or overlaps. To take one current example, the historic <y>, /j/ glide of Spanish has famously undergone fortition in Argentina, ultimately yielding [ʃ]; thus *yo* [ʃo], *yerba* [ʃerßa], *ayer* [aʃer]. This shift involves extending, strengthening, and prolonging the articulatory gesture, rather than retracting, weakening, and shortening.

Crucially, what does not follow directly from ET is productivity and abstraction. How does one produce an item for which there are no prior remembered exemplars? ET models typically propose that productivity arises from a process of 'generalization' – i.e., drawing analogies with other similar forms – which is a kind of abstraction by another name. But what about panlexical abstract operations? ET predicts that these do not occur – all phonology must be either concrete (and therefore advanced by Practice and the Feedback Loop in a frequency-dependent fashion), or analogical and therefore local and irregular (as in Bybee's 'conservation effect' which seeks to explain the preservation of high-frequency but paradigmatically irregular verb forms). These are the central challenges to an empiricist, anti-abstraction, rule-free theory like ET. Can we get all the facts right without rules, can phonology survive without a balanced diet that includes rules and abstraction along with usage?

Please cite this article in press as: Guy, G.R., Linking usage and grammar: Generative phonology, exemplar theory, and variable rules. Lingua (2013), http://dx.doi.org/10.1016/j.lingua.2012.07.007

3. The need for rules

As I see it, there is plenty of evidence that rules are still required. Many phonological processes are most straightforwardly accounted for in terms of 'rules' – which I use here to mean an abstract processing system operating over formal representations of linguistic structures, including evaluations by ordered constraints as in Optimality Theory.

One relevant example comes from child language acquisition and dialect contact. The subject (as is so often the case in child language studies) is the child of the linguist. Jesse went through earliest language acquisition in Philadelphia. At the age of 1 yr 10.5 months, he and his Philadelphia English proto-vocabulary were taken to Australia, where he promptly began spending 3–7 h each weekday in a child-care center, hearing Australian English, which, crucially for our example, does not (or did not at the time) flap intervocalic /t,d/. Hence he frequently encountered the contrast between the flapped pronunciations of words like *water, little* occurring in his until-then native vocabulary, and Aussie productions with full, aspirated voiceless stops. On an exemplarist view, one might have expected him to respond by adapting those particular lexical items. But that is not what Jesse did. Instead, beginning about 8 weeks after his arrival in Australia, he abruptly performed the most sweeping generalization possible: he apparently conjectured that the rule for translating his native vocabulary into AusEng was to DEVOICE ALL VOICED OBSTRUENTS in the appropriate position – all medial intervocalic post-tonic stops and fricatives. This abruptly generated a spectacular set of aberrant outputs, including, as one might expect, alveolar stops like *daddy* > *datty, cuddle* > *cuttle*, but also labial and velar stops like *table* > *taple, doggy* > *docky, baboo* (his assimilated baby word for 'bottle') >*bapoo, baby buggy* > *bapy bucky*. The change even encompassed fricatives: his *fuzzy bear* became his *fussy bear*, and when during this period he learned the words *driver* and *driving* from his AmEng-speaking parents, he duly pronounced them as *drifer, drifing*.

These innovative forms had, of course, no exemplars in the target variety, nor in his native speech, nor his parents' speech. Other than the words that had a flap in AmEng vs. a /t/ in Australian, he had never in his life heard anyone say anything like his aberrant pronunciations of these words. On the contrary, his Australian interlocutors provided numerous and systematic counter-exemplars against such pronunciations! The central mechanisms of ET do not explain this result, and suggest, in fact, a contrary outcome. Other avenues of explanation do not account for it either. This is not a general child language phenomenon; it cannot be a product of the immaturity of his linguistic system because he was successfully producing both voiced and voiceless intervocalic obstruents before making the shift. It is not a natural assimilation, because it goes in the direction of increased markedness; the extremely common diachronic development goes in the opposite direction, toward intervocalic voicing. The only plausible account is that Jesse performed an abstract mental operation on his lexicon, which took the form:

[-son] -> [-voice]/V[+stress]___V

(or in an OT analysis, he reranked his constraint system to achieve a comparable effect).

Interestingly, as he caught on to the fact that he had overshot the mark, Jesse began to peel back the target set for the operation of this devoicing rule by natural phonological classes, in ways that reflect his abstract underlying segmental analysis. First to go were the fricatives, so *fussy bear* went back to being *fuzzy*. Next were the non-coronal stops–velars and labials. He made no hypercorrections in this phase, no cases like *picky* > *piggy, happy* > *habby*. This is what one would expect if his mental lexicon had abstract representations in which intervocalic obstruents still had the correct underlying representations for voicing. But the flap cases were opaquely neutralized at this stage of his acquisition (correct analysis presumably depends on recognizing derivational relationships like *bat:batty*, vs. *bad:baddie*), so they were much harder to fix: he still had not correctly identified which of his intervocalic flaps were /t/s in the new dialect and which were /d/s, so he was still calling me *datty* when we added to his linguistic confusion by moving him back to the USA at age 2 yrs 6 mos. This again speaks against an exemplar model and in favor of abstract representations. All Australian exemplars of *daddy* had /d/, and all of *water* had /t/, so if his productions were purely exemplar driven, it should have been fairly straightforward to figure them out. But since he had both of these already encoded with a single underlying phoneme, these forms were more resistant to sorting out. Evidently, he was not working on exemplars of the lexical items '*water, daddy*', but on the mappings from his single flap phoneme to two different Australian English realizations.

The conclusions suggested by these facts are familiar from generative phonology: rule-based surface neutralizations of distinct URs are easy to do, while rewriting the URs to create a distinction that previously did not exist is hard to do – since there's no rule, you have to do it word by word. Exemplar theory incorrectly makes contrary predictions; since it asserts the primacy of the lexical identity in phonological operations, it implies that all operations should be 'word-by-word' in some fundamental sense, thus it fails to explain both the across the board devoicing in defiance of lexically specific counter-examples, and the absence of word-by-word differentiation of cases like *water, daddy*, that are differentiated in the exemplar cloud input but not in Jesse's productive output.

A second example, from adult speech, comes from my work with Sally Boyd on the life-span reanalysis of the morphological class of semi-weak verbs in English–forms like *keep-kept, tell-told, lose-lost, leave-left*. In Guy and Boyd (1990), as shown in Fig. 1, we presented data indicating that speakers pass through three different analyses of these

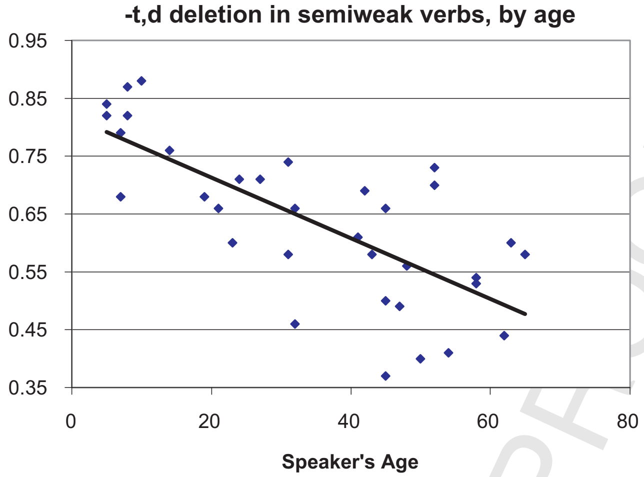

Fig. 1. Life-span morphological reanalysis: -t,d deletion in semi-weak verbs by age (from Guy and Boyd, 1990).

forms, as evidenced by their behavior in -t,d deletion. Young children have very high rates of absence of the final /t,d/ in the past tense, presumably resulting from a simple strong verb analysis of these forms: they are treated as vowel-changing rather than affixing verbs in the past tense. Thus in the figure the youngest speakers cluster at very high factor weights for deletion in this word class. In adolescence, they arrive at an irregular analysis in which these forms have underlying final stops in the past tense, but constitute unanalyzed, monomorphemic units, as shown by the fact that the young adults in the figure cluster around the deletion value of .65 that was found for monomorphemes like *old, loft*. Finally, in adult life – early middle age in many speakers – these verbs are reanalyzed again as containing a level one affix, indicated by the further decline in deletion rates among speakers over 40, showing a conservative treatment of these affixes echoing the conservative treatment of regular past tense verbs like *told, tossed*.

How does exemplar theory handle these data? Most children will hear exemplars from their parents with low deletion rates in the semiweak verbs, yet they fail to match this rate, even though they match the deletion rates in monomorphemes and regular past tense forms almost perfectly. This is demonstrated in Fig. 2, from Labov's work (p.c.) on close relatives, showing –t,d deletion rates in the various morphological contexts in one family. The seven year old son, David, closely approximates his parents' rate of deletion in monomorphemes and regular past tense forms, but deviates markedly from them in the semi-weak forms, showing instead the high deletion rates characteristic of his age group and stage of language development. David, like the other children who were studied by Guy and Boyd (1990), and by Roberts (1993), does not appear to be modeling these verbs on the exemplars provided by his parents, but on an abstract classification of them as strong, non-affixing verbs by virtue of their root vowel change. The subsequent morphological reanalyzes that give rise to the adolescent and adult systems, confined as they are to this one structurally defined morphological class, suggest further abstract analysis rather than simple usage experience. Note also that this change goes in the opposite direction to the ones predicted by the phonological process that Bybee (2002) claims underlies reduction. For Bybee,

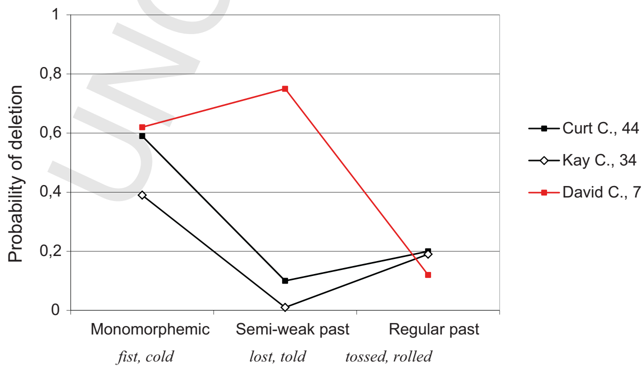

Fig. 2. Grammar and usage: -t,d deletion in one family.

Please cite this article in press as: Guy, G.R., Linking usage and grammar: Generative phonology, exemplar theory, and variable rules. Lingua (2013), http://dx.doi.org/10.1016/j.lingua.2012.07.007

lenition and deletion arises from automation of frequently repeated forms, but these forms all undergo LESS deletion as one gets older and has more experience articulating them.

Both of these examples lead to the same conclusion: speakers do not just imitate the exemplars they encounter; rather, they analyze them and translate them into mental representations consistent with the grammar they already have. Then speakers store and do operations on these mental representations, including abstract operations, and ones that reach out across the lexicon, affecting all occurrences of a segment or segment class, regardless of what words they appear in or what frequencies they display.

4. VR and the need for enriched lexical representations

So, rules (or at least, abstract operations) are one of the essential vitamins of phonological theory, and cannot be dispensed with. But as we have seen, ET does other things that MT cannot handle. What then is to be done? How can we get the best of both worlds?

One of ET's claimed advantages is its quantified, probabilistic, structure. But as noted above, the 'variable rule' (or VR) model provides another quantified alternative to MT, which retains abstract mental operations but clothes them in a probabilized framework. Let us consider how this model handles the facts.

First, the categorical properties of MT – regular sound change, categorical phonological operations, productivity of postlexical processes, etc. – are straightforwardly incorporated in VR: all categorical operations arise from a subset of the possible rules: namely, those rules with probabilities of 1. Variable outcomes have probabilities of less than one, but the model accommodates invariant processes as well. Hence all the attractive features of MT are retained in VR.

But how does VR handle the 'new facts' that ET accounts for? Probabilistic distributions of variants are of course the very thing the VR model was designed for. Similarly, gradient processes are easily modeled as probabilistic phonetic implementation rules, which get dialed up and down in extent and frequency by speakers in response to input, stylistic considerations, etc. So the main capacity of ET that is not obviously incorporated into VR are the lexical effects – lexical frequency, lexical diffusion, common contexts. What do we do about these facts?

The evidence of ET research, many papers in Laboratory Phonology, psycholinguistic research and the like makes it quite clear that lexical representations must indeed be enriched beyond the abstract phonological skeleton conventionally postulated in MT. I have argued this point in some of my own recent work (Guy, 2007, 2008, 2009; Guy et al., 2008). Notably, lexical exceptions to variable phonological processes are best accounted for by multiple underlying representations. A case in point in English is the word '*and*', which is doubly exceptional in its behavior with respect to the 'rule' of –t,d deletion. This is a lenition process, which is in fact discussed in a leading ET paper, Bybee, 2002. The word *and* is the highest frequency word eligible for final coronal stop deletion, and it ~~is~~ has the highest deletion rate. These are just the kinds of facts ET thrives upon. But VR can model these results as well, under the assumption that *and* has multiple underlying representations: one with and one without the final /d/. This is reminiscent of the ET model, but without the need to incorporate all remembered tokens of *and*. Rather, we assume each alternate representation is associated with a probability of selection. The selection of the forms without /d/ elevates the surface frequency of /d/ absence, in a manner similar to what is envisaged in ET. The double exceptionality of *and* lies in the fact that the hesitation marker '*and uh*', has an exceptionally LOW value for deletion. This suggests that *and* rather than being a single lexical item, actually participates in three underlying representations: the plain vanilla /ænd/, the underlyingly reduced /æn/ or /n/, and the discourse marker '*and uh*'.

Crucially, this model makes a quantitative prediction that does not necessarily follow from an ET analysis. Lexical exceptions like *and* that have both reduced and unreduced underlying entries will show apparently weaker surface effects of context on the rule than is the case for unexceptional, single-lexical entry words, because in a set of tokens of utterances of *and* many final /d/s are underlyingly absent, not an output of the deletion process. Hence constraints on the process do not apply to those tokens that are realizations of underlying forms that already lack the final /d/.

This prediction has been confirmed in a number of data sets in several languages. Table 1 shows the results for English for the ONZE data, with a weakened following context effect for the exceptional item. In ordinary words, following

Table 1
Contextual effects on –t,d deletion: exceptional *and* vs. unexceptional words.

Following context:	Unexceptional words			Exception: *and*	
	N	% del		N	% del
__C	1339	58.3		315	87.9
__V	1477	10.4		182	75.3
Range:	47.9%		>	12.6%	

(18 speakers from the ONZE corpus, University of Canterbury).

Please cite this article in press as: Guy, G.R., Linking usage and grammar: Generative phonology, exemplar theory, and variable rules. Lingua (2013), http://dx.doi.org/10.1016/j.lingua.2012.07.007

Table 2
Contextual constraints on –s deletion in Salvadoran Spanish: lexical exceptions vs. unexceptional words (from Hoffman, 2004).

	Non-exceptional words		Lexical exceptions (*entonces, digamos, pues*)
Following context:			
Sonorant	.60		.63
Voiced obstruent	.75		.55
Voiceless obstruent	.33		.38
Vowel	.36		.38
Pause	.44		.56
Range	.42	>	.25
Syllable stress:			
Stressed	.38		.42
Unstressed	.62		.58
Range	.24	>	.16

consonant provokes a dramatically higher deletion rate than following vowels, but for *and*, both contexts are associated with exceptionally high deletion rates, and the difference between them ('range' in Table 1) is much smaller.

A parallel case is found in Spanish final /-s/ deletion, where the discourse markers *digamos* 'let's say', *entonces* 'so/then', and *pues* 'so' have exceptionally high deletion rates, which we can attribute to the presence of additional underlying representations lacking the final /s/: *digamo, entonce, pue*. Consequently, these words should show weaker contextual effects, which they do, as shown by the ranges in Table 2, based on Hoffman's (2004) study of Salvadoran Spanish.

This model works for clear lexical exceptions like *and, entonces, digamos*. But how to get the general lexical frequency effects, and the common context effects, in VR? These appear to require further enrichment of lexical representations. A remembered exemplar cloud is one possibility–there is no fundamental incompatibility between such representations and a system that also contains rules and abstract representations. Such a hybrid system would retain all the attractive aspects of ET like social indexicality, etc. Indeed, this is the direction taken within ET by Pierrehumbert (2006). But do we need to go this far and postulate so elaborate a memory set, and an eidetic memory capacity for all human speakers that exceeds anything psychological research has ever demonstrated?

An alternative approach is to treat the whole makeup of a word as a probability function. The conventional representation of a word as a string of segments is in effect a probability function on the realization of the item, assigning a probability of 1 to the occurrence of each segment or feature. Suppose these probabilities vary with experience, in an exemplaresque fashion: hearing the word *best* occurring often, in common collocations like *bes' friend, bes' man, bes' thing*, etc., speakers assign a lower probability of full realization to the final /t/. At some point, in cases like *and* or *digamos*, this may lead to the development of new entries, with different URs, like *an', and uh, digamo*. (Such developments may be triggered by the kinds of reanalyzes that occur in grammaticalization, discourse marker formation, and lexicalization of phrasal idioms.) This permits VR to incorporate lexical effects, while retaining the abstract segmental representations that can be used to spell out novel words, translate foreign borrowings, and do things like the panlexical devoicing operation that my son so floridly performed at the age of two.

This permits an empirical test as well: the VR model predicts a distinction between lexical exceptions and words that are merely high in frequency. The effects of external constraints in high frequency words should be the same as those in low frequency words, because all deletion in such cases is ultimately accomplished by the deletion rule, and hence subject to its constraints. But when speakers create a new, exceptional entry for a word, like /*an*/ for *and*, they nullify any constraints on the rule. Hence the weakened effects of following context for *and* that we saw above. ET does not make this prediction; rather, there should be a continuously variable interaction between strength of contextual effect and frequency: higher frequency words have more deleted exemplars in the cloud which, since they already lack a final stop, are not subject to further contextually-conditioned articulatory reduction. So here's a research program: if rules are doing the work, they should show general contextual effects, but if exemplars are involved, contextual effects can vary word by word, and should, in general, get gradually weaker as lexical frequency increases.

5. Conclusions: toward a balanced diet

Exemplar theory correctly points out that MT is undernourished in certain respects, leaving it exposed to certain vitamin-deficiency diseases. These phonological analogs of scurvy and rickets include an incapacity to deal with quantitative patterns, variation and change, and lexically-specific phonetic and phonological properties. But the conclusions ET draws from these diagnoses are incorrect: the problems are not well solved by switching from phonological meat and potatos to an exemplarist macrobiotic diet. This will simply leave us exposed to other deficiency

Please cite this article in press as: Guy, G.R., Linking usage and grammar: Generative phonology, exemplar theory, and variable rules. Lingua (2013), http://dx.doi.org/10.1016/j.lingua.2012.07.007

diseases, like kwashiorkor or goiter. Specifically, a puristic ET does not do abstract operations, productivity, and the successful interpretation of forms lacking prior exemplars. In short, phonological abstraction should not be abandoned entirely just because its mainstream implementation ignores quantification. Rather, a healthy phonological theory needs to add a few nutritional supplements. Some missing vitamins are supplied by the variable rule model, and if VR is augmented by incorporating quantified lexical representations, with additional lexical entries for words with clearly exceptional behavior, this may do the rest of the job. This echoes the position taken by Pierrehumbert (2006:523): ''When we consider simultaneously the successes of Varbrul [the name of the application commonly used in VR analyses -GG] and the strong points of exemplar theory, it is clear that a hybrid model is needed.'' Significantly, VR makes testable predictions about where and how multiple entries should differ from unique representations, even if these incorporate frequency information. Testing these predictions will allow us to decide the appropriate role of rules and representations. But the important point is that phonological theory needs both the meat of grammar and rules and the vegetables of memory in order to account for all the facts, both the 'old facts' and the 'new' ones. An enriched Variable Rule model provides just this, a nutritionally sound version of phonology – a balanced diet that pays attention to quantities.

References

Bod, R., Hay, J., Jannedy, S., 2002. Probabilistic Linguistics. MIT Press, Cambridge, MA.
Boersma, P., Hayes, B., 2001. Empirical tests of the gradual learning algorithm. Linguistic Inquiry 32, 45–86.
Bybee, J., 2001. Phonology and Language Use. Cambridge University Press, Cambridge.
Bybee, J., 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. Language Variation and Change 14, 261–290.
Bybee, J., 2010. Language, Usage, and Cognition. Cambridge University Press, Cambridge.
Cedergren, H., Sankoff, D., 1973. Variable rules: performance as a statistical reflection of competence. Language 50, 333–355.
Chomsky, N., Halle, M., 1968. Sound Pattern of English. MIT Press, Cambridge, MA.
Foulkes, P., Docherty, G., 2006. The social life of phonetics. Journal of Phonetics 34, 409–438.
Guy, G.R., 1997. Violable is variable: optimality theory and linguistic variation. Language Variation and Change 9, 333–348.
Guy, G.R., 2007. Lexical exceptions in variable phonology. In: Cook, T., Evanini, K. (Eds.), Papers From NWAV 35, 13.2. Penn Working Papers in Linguistics, pp. 109–120.
Guy, Gregory R., 2008. What's in a name? The representation of lexical items in variable grammar. In: Paper Presented at the 37th Annual Conference on New Ways of Analyzing Variation (NWAV 37), Rice University.
Guy, Gregory R., 2009. Unique lexical representations or multiple exemplars? In: Paper Presented at Linguistic Society of America Annual Meeting, San Francisco.
Guy, G.R., Boyd, S., 1990. The development of a morphological class. Language Variation and Change 2 (1), 1–18.
Guy, G.R., Hay, J., Walker, A., 2008. Phonological, lexical, and frequency factors in coronal stop deletion in Early New Zealand English. In: Paper Presented at Laboratory Phonology 11, Wellington NZ.
Hoffman, M., 2004. Sounding Salvadorean, Phonological Variables in the Spanish of Salvadorean Youth in Toronto. Ph.D. dissertation, University of Toronto.
Jaberg, Karl., 1908. Sprachgeographie. Sauerländer, Aarau.
Labov, William, 1969. Contraction, deletion, and inherent variability of the English copula. Language 45, 715–762.
Paolillo, John., 2002. Analyzing Linguistic Variation: Statistical Models and Methods. CSLI Publications, Stanford, CA.
Pierrehumbert, Janet, 2001. Exemplar dynamics: word frequency, lenition, and contrast. In: Bybee, Joan, Paul Hopper, (Eds.), Frequency Effects and the Emergence of Linguistic Structure. John Benjamins, Amsterdam, pp. 137–157.
Pierrehumbert, J., 2002. Word-specific phonetics. Laboratory Phonology VII, 101–139.
Pierrehumbert, J., 2006. The next toolkit. Journal of Phonetics 34, 516–530.
Roberts, J., 1993. The Acquisition of Variable Rules: t,d Deletion and –ing Production in Pre-school Children. Ph.D. dissertation, University of Pennsylvania.
Sankoff, D., 1978. Probability and linguistic variation. Synthese 37, 217–238.
Weinreich, U., Labov, W., Herzog, M., 1968. Empirical foundations for a theory of language change. In: Winfred, P.L., Yakov, M. (Eds.), Directions for Historical Linguistics. University of Texas Press, Austin, pp. 95–195.

Please cite this article in press as: Guy, G.R., Linking usage and grammar: Generative phonology, exemplar theory, and variable rules. Lingua (2013), http://dx.doi.org/10.1016/j.lingua.2012.07.007