Lexical Exceptions in Variable Phonology

Gregory R. Guy
New York University

Lexical Exceptions in Variable Phonology

Lexical exceptions in variable phonology

Gregory R. Guy

1 Distinct Lexical Entries vs. Exception Features

Lexical exceptionality is an old problem for phonology. The general form of the problem is that languages often display apparently valid phonological generalizations that do not apply to the entire lexicon; rather, certain lexical items violate an otherwise general pattern. A well-known example from English is the vowel alternation in *serene-serenity, obscene-obscenity*, which exceptionally fails to occur in *obese-obesity*. In formal phonological theory, how are such cases to be treated?

Given the conventional architecture of the grammar—in which a phonological component models general facts about the language, while a lexicon contains the specific, contrastive facts that identify each word in the language—there are two basic alternatives for treating lexical exceptions. One is to focus on their lexically specific characteristics, and encode their distinctiveness directly into the lexicon, in the form of a distinctive lexical entry that encapsulates their difference from other words that are not exceptional. Thus in the example of *obesity*, the word could be pre-emptively listed in the lexicon with /iy/ in its underlying representation. The other approach is to focus on the general processes, and index exceptional lexical items to the phonological rules or constraints that they exceptionally observe or fail to observe. This is the approach proposed by Chomsky and Halle (1968) in SPE: words can have exception features attached to their lexical entries; these features either trigger or block the application of specific rules. The same approach has been taken in Optimality Theory in the form of lexically indexed constraints (Pater and Coetzee 2005, Ito and Mester 1999); when these are present for a particular word, they invoke different violation patterns and therefore permit different optimal candidates. A related tactic in OT is the co-phonology approach of Inkelas et al. (1997), where certain lexical items invoke distinct constraint rankings; these are targeted at lexical strata with distinctive phonologies—whole sets of words rather than individual items—but they involve the same strategy of indexing words to phonologies. (For references to treatments of variation in OT see Anttila 2002, Anttila and Cho 1998, Boersma and Hayes 2001, and Guy 1997)

The question for phonological theory is, which of these approaches is preferable, or in some sense "superior"? Which, simply, is correct? Forty years of work in formal phonology has not provided a widely accepted an-

swer; rather, phonologists dealing with lexical exceptions use either one, or even both, on an essentially ad hoc basis. But the issue should be one of concern to phonologists, because lexical exceptions raise serious questions about the nature of the mental grammar, and the two approaches imply contrasting positions about the cognitive treatment of sound systems.

One reason that no resolution to this issue has been forthcoming is precisely the focus on categorical phenomena in traditional theoretical work in phonology. When each word has only a single surface realization that always occurs, any given lexical item either categorically **is** or **is not** an exception to some general process; hence there is no possibility of interaction with other conditions that might clarify the question.

However, if we turn our attention to variable processes, a broader range of evidence becomes available that offers the prospect of an empirical test of the two approaches. Studies of phonological variation also reveal cases of lexical exceptions, but in such cases words are exceptional because they undergo a given variable process at a significantly different frequency than other phonologically comparable words, rather than by the categorical application or non-application of that process. For example, the conjunction *and* in English occurs without its final /d/ far more often than words like *hand* or *band*; hence *and* constitutes a lexical exception to the general process of coronal stop deletion. However, it is a positive exception, since it shows the process **more** often than other words.

The theoretical importance of variable lexical exceptions arises from the fact that variable processes are typically conditioned by context. Therefore, a comparison of the contextual effects on exceptional and non-exceptional items provides a diagnostic probe that can be used to quantitatively test the two theoretical treatments of lexical exceptionality that I have sketched. It should in fact be possible to empirically test what kinds of mental representations speakers are actually using to distinguish lexical items in the operations of the phonology.

Let us see why this is so: when exceptional words are distinguished in the lexicon, by their underlying form, the relevant phonological processes do not operate on them, with the consequence that such tokens will not reflect the effects of any constraints on those processes. Recall that an exceptional lexical entry pre-empts the relevant phonological processes, directly encoding the exceptional surface realization in the lexicon. But, when lexical exceptions are handled by means of exception features, they still pass through regular phonological operations, and hence should still reflect the effects of any constraints on those processes.

Thus a simple test of the two approaches is possible: we compare the effects of external contexts on exceptional and non-exceptional words; if they

both show the same quantitative pattern of constraint effects, this is evidence for an exception feature strategy. However, if the exceptional words show different conditioning from the non-exceptional, this would suggest that the exceptions are **not** undergoing the phonological processes in the same way as other words, hence favoring a lexical analysis in which the exceptions have distinct underlying representations.

2 Coronol Stop Deletion in English

This line of argumentation can be exemplified with the case of English coronal stop deletion mentioned above. How might the two approaches to lexical exceptionality account for the exceptionally high rates of final /d/ absence in the word *and*? In an exception-feature (phonological) approach, *and* is indexed with a feature triggering a raised probability of the deletion process affecting this word. In a variable rule analysis, this is implemented by associating an exception factor with this word, in a factor group in which all non-exceptional words have a residual default factor.

In the lexical approach, however, the exceptionality of *and* is captured by an alternate underlying representation lacking the final /d/: i.e., *an'*. I submit that this is what is implied by the common orthographic device of spelling *'n'* in phrases like *rock 'n' roll*. Speakers thus have two mental representations for this word which they vary between. When they select underlying *and*, it undergoes deletion at the same rate as other words. But when speakers select underlying *an'*, such tokens do not undergo deletion, but still surface lacking a final /d/. The total corpus observed on the surface is thus the sum of two different pathways to /d/ absence, yielding a higher total rate of absence.

The crucial point here is that the two approaches are not mathematically symmetrical with respect to interaction with other constraints on the process. In the exception-feature approach, all deletions are generated by the phonology, so any constraints on the relevant phonological process should be robustly evident in the output of both exceptional and non-exceptional words. In the case of coronal stop deletion, one such constraint is the strong effect of following context: more deletion before consonants than before vowels (hence more deletion in *west side* than in *west end, second son* than *second effort*). If the exceptionality of *and* is due to an exception feature indexing it for a higher rate of deletion, the following context should have the same effect, hence more deletion preconsonantally in *cheese and crackers* than prevocalically in *ham and eggs*.

However, the lexical approach makes a quite different prediction. Under this hypothesis, all tokens derived from underlying *an'* (the form without

a final /d/) never undergo the deletion process, so they should show no effect of contextual constraints. Hence one should be just as likely to say *ham 'n' eggs* as *cheese 'n' crackers*. Intuitively, I feel that this is correct. Retaining the /d/ in *ham and eggs* sounds overly precise. Of course, in this model some tokens do derive from underlying *and*, and these are affected by following context, but in the total surface corpus, the mixture of the two sources should have the effect of quantitatively weakening the effects of other constraints.

Therefore, phonological variables with lexical exceptions offer straightforward empirical tests of the two approaches. If speakers' grammars use exception features, they should have the same magnitude of constraint effects for exceptional and unexceptional words, but if their mental grammars rely on alternative lexical entries for exceptional words, these should exhibit surface attenuation of the effects of contextual constraints.

3 Quantitative Testing of Phonological Theory

As it happens, there are several phonological variables described in the literature that are known to have lexical exceptions and can provide testing sites for this theoretical problem. We begin with the case of English *and*. There is a published corpus by Neu (1980) that presents some relevant figures, repeated here in Table 1.

	Non-exceptional words		Lexical exception (*and*)	
Following context	N	% deleted	N	% deleted
obstruent consonant	572	39.3	441	95.7
vowel	495	15.8	312	82.1
Range		*23.5*		*13.6*

Table 1: Coronal stop deletion in English; following context effect and lexical exceptionality (% deleted; data from Neu 1980)

Overall, Neu finds that *and* surfaces without a /d/ some 90% of the time—an extraordinarily high figure compared with a deletion rate of about 30% for other words. The following segment effect is clear in Neu's results: for non-exceptional words the deletion rate is 39% before consonants, but less than 16% before vowels, while the exceptional case of *and* shows 95.7% deletion before consonants and 82.1% before vowels. Thus both the exception and the non-exceptional words show the effect, but the difference between them is evident in the range of values for the following context effect. If the exception feature approach is operative, these should be equal. But if the high level of /d/ absence in *and* is due to the inclusion of tokens of *an'*

that don't undergo deletion, the range of values will necessarily be reduced (see Appendix for a mathematical demonstration). And in fact, such a reduction is what we find: the range is 23.5% for ordinary words, but only about half this value, or 13.6%, for *and*. These data support the lexical selection model: there is an additional lexical entry for *and*, without a final /d/, which is selected a majority of the time.

For our second example, consider the case of final sibilant deletion in popular Brazilian Portuguese. This deletion process is also strongly conditioned by two aspects of the following context. There is an OCP place effect, and there is a manner / voicing effect, with most deletion before sonorants and least before voiceless obstruents. Crucially, the process has a significant lexical exception: the first plural verbal morpheme *–mos*: e.g., *nós fala<u>mos</u>*, *nós come<u>mos</u>* 'we speak, we eat'. In running speech, verb forms containing this agreement morpheme occur without a final /s/ at a much higher rate than non-exceptional words like *menos* 'less' and *ônibus* 'bus'. Thus in the data I collected from the VARSUL corpus covering cities in southern Brazil (cf. Zilles 2005), the deletion rate for *–mos* forms is 41%, versus only 10% deletion for other unstressed non-inflectional /s/ (in words like *menos*). Hence this process offers another test of the two approaches to lexical exceptionality. The results of separate analyses of the *–mos* forms and the non-exceptional words are given in Table 2. (Note that this table gives not raw percentages but factor weights from a multivariate analysis, which are superior in that they control for other constraints on the process.)

Features of following C		Non-exceptional words	Lexical exceptions (*-mos* forms)
Voice / Manner	sonorant	.69	.49
	voiced obstruent	.44	.58
	voiceless obstruent	.36	.44
Range		*.33*	*.14*
Place	labial	.32	.58
	coronal	.61	.53
	velar	.44	.39
Range		*.29*	*.19*
N		5880	1225
Log likelihood		-704.8	-791.5

Table 2: Final /s/ deletion in Brazilian Portuguese: following context effects and lexical exceptionality (factor weights; data from five cities in VARSUL corpus)

For non-exceptional words, the voicing / manner effect shows more deletion before voiced segments, with peak deletion before sonorants, and the place effect shows maximum deletion before coronals, confirming results found for other corpora. Both factor groups have substantial ranges: .33 for voicing / manner and .29 for place.

However, for the *–mos* words, the picture is quite different. First, both generalizations about most favorable contexts are lost. Sonorant is no longer the most favorable voicing category, and remarkably, coronal is not the most favorable place! This is striking, given the systematic evidence for OCP preferences in many phonological processes, both variable and invariant. This is strong evidence of a non-phonological process affecting these data. Second, the range of values for both factor groups is reduced in the exceptional cases, by a factor of one-third for the place effect, from .29 to .19, and by more than half for the voicing effect, from .33 to .14. The phonological effects thus appear weakened in these data, also suggesting that a pre-phonological process accounts for increased /s/ absence in *–mos* forms. The implicated "other process" is, I argue, the selection of a lexical entry like *-mo* that already lacks a final /s/.

Another way to compare the exceptional and non-exceptional cases in this analysis is to examine the log likelihood statistics—the goodness of fit measure incorporated in the Varbrul procedure. This is a negative number whose absolute value increases as the fit between model and data gets worse, and also as more tokens are added to the corpus. In these data, the non-exceptional corpus of 5900 words has a log likelihood of -705. The lexical exceptions, with a corpus only one-fifth the size (1225 words), show a **larger** log likelihood, of -792! A larger log likelihood, even with 80% fewer tokens in the analysis, means that the exceptional words fit the model much more poorly. The appropriate conclusion is that the *–mos* words are not well-predicted by purely phonological factors; something else is going on. That "something else", I suggest again, is lexical: many first plural verb forms lack a final /s/ in underlying representation, in the input to the phonology; consequently, phonological context does **not** explain their absence very well.
A third empirical test of the treatment of lexical exceptions comes from Salvadoran Spanish, which also has a final /s/ deletion process. Hoffman (2004) finds exceptional behavior in three discourse markers that show high rates of /s/ absence: *entonces* 'then', *pues* 'so', and *digamos* 'let's say'. When Hoffman analyzed these tokens separately, she found different results for the phonological constraints on the process. The results for two constraints, stress and following segment, are presented in Table 3.

		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*

Table 3: /s/ deletion in Salvadoran Spanish: stress and following context effects and lexical exceptionality. (factor weights; data from Hoffman 2004)

The following segment effect parallels the Portuguese case: vowels and voiceless consonants disfavor deletion, but voiced consonants and sonorants favor it. But the magnitude of the effect is smaller in lexical exceptions: the range of values is reduced from .42 to .25. The stress effect, in which unstressed tokens favor deletion, also shows a reduction in magnitude in exceptional words, by a factor of one-third. Recall that, the exceptional cases show a higher rate of absence overall, so it is not the case that effects are attenuated by simply a lack of evidence. Rather, these results parallel the Portuguese case, suggesting that the increased absence of final segments in the exceptional cases is due to the inclusion of items that are not conditioned by context, because they do not have the final /s/ present in their underlying representation.

4 Conclusion

All of the above examples support the same conclusion: exceptional lexical items in cases of variable phonological processes are best treated lexically, by means of alternative underlying representations, rather than by exception features. If the exception feature treatment were valid, at least some case ought to show a constant effect of other phonological constraints across exceptional and non-exceptional words. But this is not the case. Five different constraint effects on three variable processes in three different languages all exhibit a reduction in the magnitude of the effect in exceptional lexical

items. Thus I find no evidence at all suggesting that exception features operate in the mental grammars governing these cases of phonological variation.

These facts suggest a further prediction. In principle, the exception feature approach permits both positive and negative exceptions—that is, there should be words that undergo phonological processes at both exceptionally high and exceptionally low rates. The exception feature in a variable rule analysis is a factor weight: it could be above .5 in value, promoting application, or below, inhibiting application. But the lexical approach, which encodes outcomes directly in the lexicon, does not permit exceptionally low rates of occurrence, at least for deletion processes. There is no reasonable way to construct alternative entries for *and*, *–mos*, or *entonces* that will resist the deletion processes more than other words. Hence the lexical approach predicts that only words with exceptionally high rates of occurrence should be found. It is my impression that this prediction is consistent with the cases discussed in the literature.

Assuming this prediction is also confirmed, we have strong quantitative evidence bearing on the theoretical issue at hand: all the data are consistent with the predictions of the lexical approach to exceptionality, and none are consistent with an exception-feature treatment. Speakers encode lexical exceptionality by means of alternative underlying representations, rather than by twiddling the phonology. This is a potentially decisive resolution based on variable phonology that has not been achieved in four decades of formal work on categorical processes.

Appendix

The mathematics of range reduction in lexical exceptions

1) Consider a variable phonological process P that is contextually conditioned so that it applies less often in some disfavorable context A and more often in some favorable context B. There is a lexical exception to this process which has two URs, one of which is unrestructured, and undergoes P at the same rate as other lexical items (e.g. *and* = [and]), and another of which is restructured so that it encodes the effects of P directly in the UR and hence does not undergo P (e.g. *and* = [an]). In production, the unrestructured UR is selected with a frequency *f*, while the restructured UR is selected with frequency 1–*f*.

2) In a multiplicative model of constraint effects, we can define the following values:

Let *a* represent the application rate process P, in the less favorable context A.
Let *c* represent the constant (where $c > 1$) by which the application rate is increased in the more favorable environment B.

3) In unexceptional lexical items, in environment A the application rate is *a*, and in B it is *ca*. The range of these values is: $ca - a$

4) For the exceptional lexical item, the set of apparent (surface) "applications" of P will actually represent the sum of the restructured URs plus those unrestructured URs that also undergo P. In environment A, for a corpus of N tokens, these will be:

restructured URs = $(1 - f)N$
unrestructured URs that also undergo application = afN

The sum of these is

$$(1 - f)N + afN = (1 - f + af)N$$

Dividing by N to get the apparent application rate gives

$$1 - f + af, \text{ or } 1 + (a - 1)f$$

In environment B, the surface applications will be:

restructured URs = $(1 - f)$N
unrestructured URs that also undergo application = $cafN$

The sum of these is:

$$(1 - f)\text{N} + caf\text{N} = (1 - f + caf)\text{N}$$

Dividing by N to get the apparent application rate gives:

$$(1 - f + caf) = 1 + caf - f = 1 + (ca - 1)f$$

The range of these values is (the value for context B minus the value for context A), i.e.:

$$(1 + (ca - 1)f) - (1 + (a - 1)f) = 1 + (ca - 1)f - 1 - (a - 1)f = (ca - a)f$$

5) Comparing the exceptional and unexceptional items, the ranges are:

unexceptional items: $(ca - a)$
exceptional items: $(ca - a)f$

Thus, the range for exceptional items equals the range for unexceptional items multiplied by the frequency of lexical selection of the unrestructured underlying representation. Hence whenever $f < 1$, i.e. whenever a lexical item has a restructured UR selected that does not undergo the process P, the contextual factors affecting P must show a range of values that is smaller than for lexical items that have no such exceptional URs.

References

Anttila, Arto, 2002. Variation in phonological theory. In *The handbook of language variation and change*, ed. J.K. Chambers, P. Trudgill and N. Schilling-Estes, 206–243. Oxford: Blackwell.

Anttila, Arto and Young-mee Yu Cho. 1998. Variation and change in Optimality Theory. *Lingua* 104:31–56.

Boersma, Paul and Bruce Hayes, 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32:45–86.

Chomsky, Noam and Morris Halle, 1968. *The sound pattern of English*. New York: Harper.

Guy, Gregory R. 1997. Violable is variable: Optimality theory and linguistic variation. *Language Variation and Change* 9:333–348.

Hoffman, Michol, 2004. *Sounding Salvadorean, Phonological variables in the Spanish of Salvadorean youth in Toronto*. Doctoral dissertation, University of Toronto.

Inkelas, Sharon, Cemil Orhan Orgun & Cheryl Zoll. 1997. Implications of lexical exceptions for the nature of grammar. In *Constraints and Derivations in Phonology*, ed. Iggy Roca, 393–418. Oxford: Clarendon.

Ito, Junko and Armin Mester. 1999. The phonological lexicon. In *The handbook of Japanese linguistics*, ed. N. Tsujimura, 62–100. Malden, MA: Blackwell.

Neu, Helene, 1980. Ranking of constraints on –t,d deletion in American English. In *Locating language in time and space*, ed. William Labov, 37–54. New York: Academic.

Pater, Joe and Andries Coetzee, 2005. Lexically specific constraints: Gradience, learnability, and perception. In *Proceedings of the 3rd Seoul International Conference on Phonology*, 85–119. Seoul: The Phonology-Morphology Circle of Korea.

Zilles, Ana M. S., ed. 2005. *Estudos de variação lingüística no Brasil e no Cone Sul*. Porto Alegre: Editora UFRGS.

Department of Linguistics
New York University
719 Broadway, #452
New York, NY 10003
gregory.guy@nyu.edu