1-1-1994

Variation and Optimality

Bill Reynolds  
*University of Pennsylvania*

Hadass Sheffer  
*Swarthmore College*

This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/pwpl/vol1/iss1/7](http://repository.upenn.edu/pwpl/vol1/iss1/7)  
For more information, please contact libraryrepository@pobox.upenn.edu.
Variation and Optimality
Variation and Optimality
Bill Reynolds and Hadass Sheffer

1 Introduction

This study proceeds from two fundamental assumptions which have nevertheless not always received wide acceptance in the fields of phonology and sociolinguistics: (1) that variable or optional rules, and the empirical, quantitative study of variation in language, may have important implications for theories of phonological/morphological derivation; and (2) that a thorough understanding of current phonological theory can be of great importance in interpreting sociolinguistic data and in elevating such interpretations beyond the merely descriptively adequate to the explanatory.

Sociolinguistic studies have long concerned themselves primarily with a descriptive view of sound change and variation phenomena. Most recently — based in large part on the pioneering work of Labov, Sankoff, and others — such studies have taken the form of quantitative analyses, whereby a variable rule would be examined for the purpose of determining its rate of application and the phonological environments and/or sociological factors which tend to trigger or suppress it. These probabilistic analyses of variable rules allow us to characterize the relative weight played by the various factors (linguistic, stylistic, social, etc.) which influence — consciously or otherwise — the speaker’s selection of one of two or more alternative forms in speech production. Such variation is seen to play a key role in both the process of diachronic change and the interaction of human beings in society. Furthermore, the claim is made that the empirical study of variation in production can tell us as much as, if not more than, intuitive judgments of grammaticality with regard to the speaker’s linguistic competence.

In phonology, however, as well as in other areas of formal linguistics, variable rules (or “optional” rules, as they are generally referred to) are too often totally ignored, or their importance at least not fully acknowledged. It is not that phonologists fail to appreciate the theoretical implications to be derived from empirical linguistic research. Rather, variable rules tend to be dismissed as inherently uninteresting for phonology. In general, they are regarded as being primarily not only questions of style and social stratification but also as belonging properly within the domain of phonetics. Yet it is far from clear that performance rules per se are of no interest to formal linguistics. As Labov has noted,

one may set aside variable rules on the ground that they are rules of performance. The less said about this “wastebasket” use of the performance concept the better. For it must be noted that the great majority of our transformational and phonological rules may also be characterized as “performance” rules....

(1972, p. 226.)
If phonologists display a tendency to dismiss out of hand the relevance of sociolinguistic studies of variation, it must be stated that there is an equally strong tendency for sociolinguists to disregard the implications of current developments in phonology for their own field. Since the publication of Chomsky and Halle's *The Sound Pattern of English* in 1968, there has been a virtual explosion in phonological theory in many different directions, resulting in a large degree of modularization within the field. The development of autosegmental phonology has led to a multi-tiered (rather than linear) approach to the representation of both segmental and suprasegmental elements. Many changes have taken place in the way we view the organization of the lexicon and the interaction of phonological and morphological rules. Competing theories of the nature of feature specification of underlying forms have provided new ways of dealing with the question of abstractness. In morphology, the postulation of a prosodic hierarchy has had a profound effect on the way we view aspects such as syllable structure, affixation and reduplication processes, etc. And, within the last three years, the role played by constraints on representations has attained increasing prominence as a means of expressing the relationship between underlying and surface forms.

Yet sociolinguists often feel — rightly or wrongly — that these developments in phonological theory have little relevance to their own concerns. This is in large part attributable to the general direction of such developments away from a theory of rules and toward a theory of representations. Further, this direction in phonology has brought with it a diminishing of interest in the question of extrinsic rule ordering; instead, there is now general agreement that the order of rules should be seen to fall out from considerations of rule types (e.g., phonological vs. morphological, lexical vs. postlexical); the levels at which they apply; and, most recently, the ranking of “constraints on well-formedness of linguistic representations” (McCarthy & Prince 1993, p. 1).

2 Optimalit y Theor y

In this paper, then, we explore ways in which one particular approach to formal phonology — namely, the constraint ranking approach embodied in Optimality Theory, as developed by Alan Prince and Paul Smolensky (1991a, b, 1992, 1993) and advanced by John McCarthy (McCarthy 1992, 1993; McCarthy & Prince 1992, 1993) — might have relevance for sociolinguistic studies of variation. No theoretical approach, of course, can take the place of quantitative analysis for purposes of determining the complex interplay of sociological factors like age, sex, education, socioeconomic status of both speaker and addressee, and the context of the situation — factors which help to determine the choice a speaker makes with regard to the use of a given variant. Rather, through Optimality Theory we can address phonological processes themselves, and the phonological environments which trigger or suppress them, in a more systematic, abstract, and formalized way. In other words, how can we explain the linguistic factors which condition documented cases of phonological variation using the insights provided by a particular phonological theory?

This paper addresses precisely that question. Using Optimality Theory, we will examine a well-documented sociolinguistic variable and see how the processes involved can be explained by universal tendencies of language.

Linguistic variation presents itself as differing yet related forms. Generative phonology has treated these as several alternative surface forms derived from one underlying form, which itself may never actually surface. Following Optimality Theory, we would like to suggest that rather than actually deriving the variant surface forms from a single, abstract underlying form, speakers choose directly from a range of possible output forms. Optimality Theory provides a new perspective on the generation of these surface or output forms which lends itself particularly well to this kind of variation data. We
Variation and Optimality

Reynolds and Sheffer

claim that this is a much more realistic approach to the analysis of variation in natural language than one which depends on positing different derivational rules and/or rule orderings. In addition, Optimality Theory allows us to consider every possible surface form for a particular input at once, even when there are many interacting factors.

According to McCarthy (pers. comm.), “variation is a consequence of not-ranking of constraints that in principle conflict. The optimal form, then, is non-unique.” The choice between these forms falls to sociological factors. In this approach, there is one ranking hierarchy with a choice between equally well-formed candidates. An alternative explanation, proposed by Kiparsky (1993), entails a stricter view of the ranking hierarchy, under which all competing constraints must be ranked. Thus, for each ranking hierarchy there can be only one winning form. However, a speaker may have more than one ranking hierarchy. In such a case, each hierarchy will produce one winning form. The speaker then chooses between the ranking hierarchies.

The difference between these two approaches may seem at first glance to be purely semantic. However, since according to Optimality Theory a grammar is defined as a ranked hierarchy of constraints, Kiparsky’s approach necessarily implies the simultaneous existence of separate, competing grammars within the mind of each individual speaker — one complete grammar, in fact, for each instance of variation within that speaker’s language. In McCarthy’s alternative, on the other hand, it is not clear how one would allow for the conditioning of variation by phonological environments; yet such conditioning is well-documented in any number of sociolinguistic analyses. Further, neither of these approaches seems well-equipped to capture the gradience which characterizes many variable processes.

For further discussion of these issues (and suggestions for alternative approaches), we refer the reader to Liberman (1994), Reynolds and Nagy (1994), and Reynolds (to appear). For the present, for ease of representation we will rank the constraints showing variation, without thereby committing ourselves firmly to the position that the resulting rankings necessarily reflect separate, competing grammars.

3 Spanish /s/

One of the most well-studied variables in the entire field of sociolinguistics is /s/ aspiration/deletion in Caribbean Spanish. This is a lenition process whereby syllable-final /s/ may surface alternately as [s], [h] or φ. /s/ deletion, however, also provides a classic instance of a situation where the factor groups which favor or disfavor its application are well-attested, yet the phonological structure which underlies the process has received little if any attention. By approaching the problem from the standpoint of Optimality Theory, however, we can show how deletion (or lenition) of syllable-final /s/ falls out very naturally from considerations of syllable structure and licensing constraints in Spanish.

Optimality Theory provides formulations of universal constraints that account for these tendencies of language:

1. PARSE: Every element of the phonological representation must be dominated by, or associated to, an appropriate node of the prosodic tree.
2. FILL: Every node of the prosodic tree must dominate, or be associated to, something.
3. ONSET: Onset position is optimally filled.
4. EDGEMOST(x; E; D): Item x is situated at edge E of domain D. This constraint penalizes forms according to the extent that the item x is removed from the edge E.
These constraints will be used in the analysis below.

A crucial phonological conditioning factor for the process of syllable-final /s/ deletion in Spanish — as almost all analyses have shown — is following environment: That is, deletion is most favored before a pause; less before a following consonant; and least of all before a vowel, where it can easily be syllabified into the onset of the following syllable. In other words, given a CVCV sequence, Spanish, like most other languages, will syllabify it as CV.CV, rather than CVC.V. It is a universal generalization about languages that they prefer onsets to codas. A coda /s/ that coincides with a word boundary can thus syllabify into the onset of the first syllable of the following word, provided this following word is vowel-initial. If it begins with an initial consonant, there is a much stronger tendency to block this type of syllabification (and hence /s/ deletion), especially if this initial consonant is less sonorant than the /s/. This is due to two universal constraints which call for rising sonority in syllable-initial consonant clusters (i.e., in onsets) and falling sonority in codas. We may call these constraints ONSRISE and CODFALL:

(5) ONSRISE: If there is more than one position in an onset, these positions must be filled by segments which are rising (toward the nucleus) in sonority.

(6) CODFALL: If there is more than one position in a coda, these positions must be filled by segments which are falling (away from the nucleus) in sonority.

It is the ONSRISE constraint in Spanish which accounts for the word-initial epenthetic vowel in pairs such as Lat. statio, stationis > Sp. estación and Lat. schola > Sp. escuela, for example. In the Latin forms, the initial /s/ violates ONSRISE; in Spanish, where this constraint is ranked higher (and thus resists violation), addition of a word-initial /e/ allows the /s/ to be part of a coda rather than an onset. (It might, of course, be pointed out that this same epenthetic vowel is also found before s+Nasal and sl onset clusters — that is, in conditions where ONSRISE would not be violated. However, words with these clusters are invariably borrowings, and thus the epenthesis in such forms can safely be assumed to be due to analogy.) Finally, in the case of a pause (i.e., when /s/ occurs at the end of an intonational phrase), there is the strongest possible incentive to delete the /s/, since there is no possibility of its syllabification into a following onset.

Spanish has long differentiated itself from its Latin roots by the addition of a very common cross-linguistic constraint on syllable coda consonants. This is a rephrasing of the weaker licensing that many languages associate with a coda position (as has been noted many times in the literature; see, e.g., Steriade 1982; Ito 1986, 1989; Yip 1991). Prince and Smolensky have termed this the Coda Condition (1993:99):

(7) CODACOND: A coda consonant can have only coronal place or else no place specification of its own at all.

Specifically, consonants which are licensed by the coda position in Spanish consist solely of the coronals /r, l, s, d/ (coronal being the default place of articulation and thus unspecified in underlying representation; see, e.g., Paradis and Prunet 1991, among others) and nasals which are homorganic with a following (onset) consonant.

We here advance the theory that CODACOND, in many dialects of Caribbean and Central American Spanish, has over time been replaced by a different constraint, which Prince and Smolensky term the Possible Coda Parameter (1993, p. 159):

(8) POSSCOD: The possible codas are those segments with sonority value greater than or equal to a cutoff parameter $\pi_{\text{Cod}}$. 

106
That is, over the years — quite likely due to the fact that the large majority of possible coda consonants are [+sonorant] — succeeding generations of Spanish speakers have reinterpreted the constraint on codas as one involving sonorancy rather than place of articulation. The cutoff point for many such speakers has become variable, so that under certain conditions it may be a fricative (/s/), while under others it has moved up a notch along the sonority hierarchy to a nasal.

Final /s/ may also surface not only as [s] or ø (i.e., the root node is not linked to a mora), but as /h/ (a sort of in-between variant, where the place node is not linked to the root). In order to account for this fact, we must posit a variably ranked independent constraint on /h/ in coda position. Such a constraint is not uncommon (it applies, obviously, in English); the question is how to express it in such a way that would have universal application.

McCarthy (1993) provides a way to approach this problem, by “exploding” or subdividing the PARSE constraint, as follows. The candidates which GEN generates are assumed to consist of a segmental string with its associated moras. Segments themselves consist of feature sets in underlying representation; these features may be parsed or not in the candidates emitted by GEN. We can refer, then, to a subfamily of Parse constraints, such that the root node is parsed by the mora; the laryngeal and place nodes by the root node; etc. For our purposes here, the following two Parse constraints will be crucially ranked:

\[ (9) \text{PARSE-RN: The root node is parsed.} \]
\[ (10) \text{PARSE-PN: The place node is parsed.} \]

Thus, in the case of a word such as \textit{restar} ‘to deduct or reduce’, the optimal surface form will depend both on the cutoff parameter for POSSCOD and the relative ranking of the constraints PARSE-PN and PARSE-RN. If the POSSCOD parameter is a fricative, then the candidate \textit{res.tar} will emerge the winner (regardless of how PARSE-PN and PARSE-RN are ranked):

\[ (11) \text{ONSRISE, FILL, EDGE >> POSSCOD} \]

<table>
<thead>
<tr>
<th>Candidates</th>
<th>ONSRISE</th>
<th>FILL</th>
<th>EDGE</th>
<th>POSSCOD</th>
<th>PARSE-PN</th>
<th>PARSE-RN</th>
</tr>
</thead>
<tbody>
<tr>
<td>res.tar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reh.tar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>re&lt;s&gt;.tar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>re.star</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>re.s&lt;t&gt;ar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*!</td>
</tr>
<tr>
<td>re.sD.tar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>re.sD.tD.Dr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

On the other hand, if the possible coda parameter cutoff point is higher in sonority than /s/, the winning candidate will depend entirely on the relative ranking of PARSE-PN and PARSE-RN:
(12)  $$\text{POSSCOD} >> \text{PARSE-PN} >> \text{PARSE-RN}$$

<table>
<thead>
<tr>
<th>Candidates</th>
<th>POSSCOD</th>
<th>PARSE-PN</th>
<th>PARSE-RN</th>
</tr>
</thead>
<tbody>
<tr>
<td>res.tar</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reh.tar</td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>re&lt;s&gt;.tar</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

(Note that the candidate re<s>.tar does not violate PARSE-PN; the place node is still associated with its appropriate mother node (i.e., the root node), but whether or not the root node is itself parsed is entirely irrelevant.)

(13)  $$\text{POSSCOD} >> \text{PARSE-RN} >> \text{PARSE-PN}$$

<table>
<thead>
<tr>
<th>Candidates</th>
<th>POSSCOD</th>
<th>PARSE-RN</th>
<th>PARSE-PN</th>
</tr>
</thead>
<tbody>
<tr>
<td>res.tar</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>re&lt;s&gt;.tar</td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>reh.tar</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

An advantage of the competing grammars approach advocated by Kiparsky is that we may also thereby account for what is perhaps the strongest factor favoring or disfavoring /s/ deletion — namely, the so-called parallel processing or serial effect (see, e.g., Scherre and Naro 1992). If intra-speaker variation (as represented by differently ordered rankings of particular constraints or different cutoff parameters within the speech of a single individual) is a reflection of the fact that speakers have different grammars in their heads, then we may make a very apt analogy with the case of code-switching. The serial effect states that in the case of, e.g., an NP consisting of Det + Adj + N, the greatest determinant of whether plural /s/ appears on the Adj and N is whether it appears on the Det. This is hardly surprising, given the relative unlikelihood of switching grammars in the middle of an NP.

Finally, our theory predicts that those dialects in which /s/ deletion is prevalent will have an even stronger tendency to delete word-final /d/. And in fact, through personal communications with Carmen Richardson and Tom Morton of the Penn Linguistics Department, we have learned that final /d/ is deleted almost categorically in these dialects in words such as sed [se] ‘thirst’, usted [u-ste] or [u-te] ‘you’, and verdad [βer-dα] ‘truth’. In his examination of Spanish as it is spoken in the Americas, del Rosario (1970) reports that word-final stops (i.e., /d/) typically do not get pronounced, and that this loss occurs with greater frequency in the Caribbean, Central America, Chile, and the interior and south of Spain.

4 Conclusion

We have shown how Optimality Theory neatly describes the phonological structure underlying certain variable rules. In addition, this approach allows us to actually pinpoint
the constraints that are causing the variation, and consider the various possible output forms. Furthermore, subsequent ranking (or re-ranking) of these particular constraints gives us a basis for speaking of a change in the grammar of a particular speech community (a variant which has taken hold to the extent that it is now categorical). This is analogous to the view first described by Halle (1962), and corroborated by Kiparsky (1968), that "two dialects or two stages in the historical development of a language could have the same underlying representations and the same rules, but differ simply in terms of the ordering of the rules" (Kenstowicz 1994, p. 98).

References


