

Representations, Derivations, Constraints: The Theory of Vowel Harmony

Marc van Oostendorp

1 Background

- ☞ Many of the changes in generative phonology after *The Sound Pattern of English* have been driven by considerations of languages different from English. In this class, we study one specific phenomenon, found in many languages but not in (any Standard variety of) English.
- ☞ Many developments can be seen as an interactions of considerations about representations and considerations about representations
- ☞ If a language L has vowel harmony, this means that the vowels in L can be subdivided into disjoint subsets $S_1 \dots S_n$ (called harmonic sets), such that all the vowels within a certain domain (e.g. a word) belong to a harmonic set (quasi-formally: \forall word $w \in L_{\mathcal{L}} \forall$ vowel $v \in w: \exists S_i \in L_{\mathcal{L}}, 1 \leq i \leq n: v \in S_i$).
- ☞ For instance: Tangle, a Chadic language spoken in Nigeria, has two subsets of vowels: /i, u, e, o/ and /ɪ, ʊ, ε, ɔ, a/, and words such as the following:

- (1) a. $\eta\upsilon l d \epsilon d \epsilon$ 'dog'
- b. $se b - u$ 'look (imp.)'
- c. $k \epsilon n - \upsilon$ 'enter (imp.)'
- d. $t u g - o$ 'pounding'
- e. $w \upsilon d - \circ$ 'farming'

Words like $*\eta\upsilon l d \epsilon d \epsilon$ or $*se b - \upsilon$ are typically absent from the language. The behaviour of the suffixes shows that harmony is an active process in the language.

- ☞ The existence of vowel harmony in natural language raises many questions about the nature of the computational process involved, such as:

- ✿ why does harmony often run from stems to affixes but never the other way around?
- ✿ why is there no pervasive process of consonantal harmony alongside vowel harmony?
- ✿ how should we understand opaque and transparent vowels?

We will divide the history of generative grammar in three periods in which the emphasis was on (i) rules, (ii) representations and (iii) constraints, and see that each of these concentrated on certain questions, leaving aside others.

2 Rules and Exceptions

The first question that arises if we study vowel harmony from the point of view of SPE phonology, is: what is the nature of the rule involved? There are at least two candidates ([+F] and [-F] characterize harmonic sets; [X] is a morphological feature):

- (2) a. phonological context: $V \rightarrow \left[\begin{array}{c} +F \\ V \end{array} \right] / \left[\begin{array}{c} +F \\ V \end{array} \right] C_0 \text{ ---}$
- b. morphological context: $V \rightarrow \left[\begin{array}{c} +F \\ V \end{array} \right] / \text{---}]_X$

There are several possible objections against (2a). One is that it cannot explain the existence of so-called neutral vowels, e.g. in Hungarian. In this language one of the relevant features is [\pm back] (3a). However, the vowels *i*, *í* and *é* (the accent denotes length in Hungarian orthography) are ‘neutral’, this means that they can coöccur with either a front or a back vowel, subject to lexical specification:

- (3) *Hungarian*
- | | |
|---|--|
| <p>a. <i>Back vowels</i></p> <p>fog-unk ‘tooth POSS 1SG’</p> <p>lúd-nak ‘goose DAT’</p> <p>kalap-ból ‘hat ELAT’</p> | <p><i>Front vowels</i></p> <p>bőr-ünk ‘skin POSS 1SG’</p> <p>tök-nek ‘pumpkin DAT’</p> <p>kürt-ből ‘horn ELAT’</p> |
|---|--|
- b. *hég-am* ‘my rind’ (**hég-em*), *kés-em* ‘my knife’ (**kés-am*)

One possibility — mentioned by Paul Kiparsky in his 1968 article ‘How abstract is phonology?’ (this article appeared as Kiparsky, 1982)— is to assume that *hég* has an underlying back vowel, e.g. schwa ($h\text{ə}j$). Kiparsky rejects this proposal, because he thinks it is ‘too abstract’: there is no obvious reason why we would choose ə rather one of the many other non-front vowels that Hungarian happens not to possess.

An alternative would be an analysis along the lines of (2b), with a purely morphological diacritic; this was indeed the first analysis of generative analysis of vowel harmony, the *root-marker* theory of Lightner (1965). All 'normal' vowels within the domain of, say [+X] would receive a specification [+back], while all the vowels within the domain of [-X] would receive a specification [-back]. Yet to *i, í, é* another rule would apply, turning them to [-back] even in the domain of [+X]. This approach is obviously criticized (by Kiparsky, but also in a much more recent textbook, Kenstowicz (1994)) because of its use of the mysterious feature [±X].

This type of solution was quite popular in 'Natural Generative Phonology' a non-mainstream school of thought in (mainly) the 1970s (Hooper 1976). Some varieties of Spanish, such as the dialect of Granada, have the following pattern:

(4) *Granada Spanish* ([±tense] Harmony)

Orthography	Singular	Plural	Gloss
<i>pedazo</i>	[peðaθo]	[peðaθo]	'piece'
<i>alto</i>	[alto]	[alto]	'tall'
<i>cabeza</i>	[kaβeθa]	[kaβeθa]	'head'
<i>selva</i>	[selva]	[selva]	'forest'

The difference between singular and plural nouns is thus expressed as a difference between all tense vowels or all lax vowels within the word. The history of this is that there has been a plural suffix -s, as there still is in many other dialects of Spanish, closing the final syllable. Closed syllables are often lax in Spanish, Southern French, Germanic languages, etc. This laxing process has become independent after the suffix s was lost. According to Hooper, it spread through the word since the difference between tense and lax vowels is difficult to hear (yes, this is a functional explanation).

(5) $V \rightarrow [-\text{tense}] / X _ _ X\#]_{N,A;+\text{plural}}$

A severe problem for this purely morphological approach is the existence of disharmonic roots. In many languages with vowel harmony, roots can be disharmonic; but even in these cases, a harmonic feature spreads to affixes:

(6) a. *Akan* (ATR Harmony) (e is a lax e)

<i>Disharmonic roots</i>	<i>Harmony with suffix</i>
<i>bisa</i> 'ask'	<i>o-bisa-i</i> 'he asked it'
<i>kari</i> 'weigh'	<i>o-kari-i</i> 'he weighed it'
<i>ñinsen</i> 'be pregnant'	<i>o-ñinsen-i</i> 'she became pregnant'

b. *Turkish* ([±back] Harmony)

<i>Disharmonic roots</i>	<i>Harmony with suffix</i>
<i>bobin</i> 'spool'	<i>bobin-ler</i> 'spools'
<i>pilot</i> 'pilot'	<i>pilot-lar</i> 'pilots'
<i>kitap</i> 'book'	<i>kitap-lar</i> 'books'

This behaviour is unexpected in an analysis based on morphological diacritics, and therefore this theory fails. Kiparsky proposes an alternative in which harmony within the root and across the stem-affix boundary are separate phenomena, but obviously this is not a very attractive solution, conceptually. It even faces empirical problems in light of the fact that there are affixes which display a similar behaviour. We can find these in Turkish, for instance:

- (7) *Turkish* ([±back] Harmony)
- a. *üç* 'three', *üç-gen-ler* 'triangles'
 - b. *altı* 'six', *altı-gen-ler* 'hexagonals'
 - c. *çok* 'many', *çok-gen-ler* 'polygonals'

Another aspect of Kiparsky's proposal was that it gave a combination of phonological and morphological features. Thus the Hungarian words in (3b) were represented as /he:j/ and /ke:s/, respectively, but the former carried a morphological feature [-Vowel Harmony].

One problem with this analysis was that it was not able to explain why it is only words with one of the vowels *i*, *í*, *é* that can have a morphological marking. But even the facts in (6b) are not without problems. Another issue is that the neutral vowels are not subject to harmony, but harmony seems to skip them: if a polysyllabic word ends in a neutral vowel, a following suffix harmonizes to the vowel preceding the neutral vowel:

- (8) *papír-nak* 'paper+DAT', *kávè-nál* 'coffee+ADESS'

We can only solve this by assigning the 'morphological' feature [-Vowel Harmony] to the neutral vowel specifically. But this makes the analysis very murky; the notion of a 'morphological feature' has now become completely technical.

Thus also this third rule-based approach faces severe problems. Since morphological and phonological features are the only elements that are available for phonological computation in this model, there does not seem to be a way out given the assumptions of the SPE: the basic properties of vowel harmony (a process that is attested in many, genetically unrelated, languages of the world) cannot be described in a satisfactory way. Rules simply do not have the force to express the necessary generalisations. And we cannot even hope to begin to analyse properties such as the fact that vowel harmony involves *vowels*.

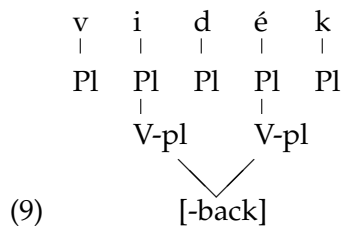
3 Representations

Many authors have noticed that the problem with the rule-based approaches just discussed is that they do not have a sufficiently sophisticated theory of representations: we need to be able to express that the phonological, harmonizing feature which is responsible for harmony belongs to the whole word, not just to an individual segment. Indeed, vowel harmony was among the primary motivations for the introduction of autosegmental and metrical models of phonology.

It should be noted that ideas about more sophisticated models of phonological representation were already present in some schools of (structuralist and other pregenerative thought). One name which is often mentioned in this connection is Firth.

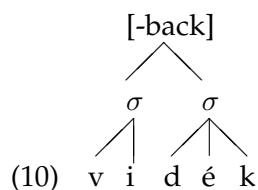
Some of the questions stated above can be answered using the mechanism of three-dimensional phonology. Let us start with the last question: why can vowel harmony skip consonants but can't consonant assimilation skip vowels? Actually, in this case, two answers are available, one from autosegmental phonology and one from metrical phonology.

In the former case, the harmonic feature is *lower* than the segment. It is assumed that consonants and vowels have the following representation for place features respectively:



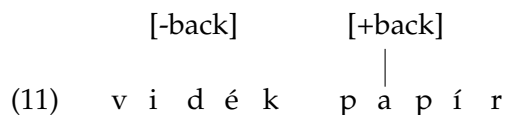
All segments have a place feature, but only vowels (and consonant with a secondary place of articulation) have a V-place feature. Assuming that spreading is strictly local, place spreading cannot take place between consonants, but it can between vowels, which are visible for one another over a longer span. (One of the nice features of this is that it can actually explain some of the interaction with complex consonants in Turkish, but I cannot show this here). A severe problem with this approach is that it is stipulative to a large extent: there is not a lot of independent evidence that vowels have this extra structure.

In the latter case, it is assumed that the harmonic feature is *higher* than the segment. It does not connect to the segment directly, but rather to the syllables dominating them:



The feature percolates to the head of the syllable, i.e. to the vowel. (A potential problem for this approach is that consonants do sometimes interact with vowel harmony.) It is obviously undesirable to have two separate ways of representing the same phenomena, and this shows one of the disadvantages of a representational approach: the structures tend to become baroque and capable of representing virtually everything.

Other problems seem still harder. Let us take a look at Hungarian (back) harmony. Assume that $[\pm\text{back}]$ is an autosegment, based on a separate tier. (Harmonic stems like *vidék* 'countryside' have a floating feature, disharmonic stems like *papír* 'paper' and *kávè* 'coffee' have a feature which is preattached to some segment.)



There is no need for a diacritic morphological feature anymore, since phonological features can function as such. The Granada case gets an elegant account, since the plural 'suffix' now consists of a floating feature $[-\text{tense}]$. Notice, however, that this means that we still have many of the same problems as the 'morphological' approach above.

One problem is that there still seem to be two processes of harmony: one within the harmonic root (a floating feature getting linked to all vowels simultaneously) and one from the root to an affix. This is actually formally almost the same process in an autosegmental framework (linking a feature to a host), however we still need to take into account why features do not spread in disharmonic roots such as *papír* (and similar questions hold for Turkish, as we have seen). We will return to this problem below.

Another issue is the behaviour of the two neutral vowels. How could these be represented? Various proposals were made during the heydays of autosegmental phonology. The key observation is that /i, í/ and possibly /é/ do not have a harmonic counterpart: there is no high (or mid) unrounded back vowel in Hungarian. In Turkish, there is (viz. /ɨ/), and this language does not have neutral vowels of the type that Hungarian does. There should be a connection between these two types.

Importantly, the most interesting approach (and for some time, the standard approach) to this problem seems to depend more on derivations than on representations. Assuming $[\text{+back}]$ to be the active phonological feature,

we could assume that (i) spreading applies first, across the board and (ii) afterwards, there is a rule delinking the feature [+back] from all the neutral vowels (remember that this can be done since the vowels in question do not contrast in the front/back dimension):

$$(12) \text{ Clean-up rule: } \begin{bmatrix} -\text{low} \\ -\text{round} \\ +\text{back} \end{bmatrix} \rightarrow \{\text{delink } [+back]\}$$

$$(13) \begin{array}{c} [+back] \\ | \quad \diagdown \\ p \ a \ p \ \acute{i} \ r \ n \ a \ k \end{array} \Rightarrow \begin{array}{c} [+back] \\ | \quad \diagdown \\ p \ a \ p \ \acute{i} \ r \ n \ a \ k \end{array}$$

Alternatively, we may assume that [-back] is the active feature:

$$(14) \text{ Clean-up rule: } \begin{bmatrix} -\text{low} \\ -\text{round} \end{bmatrix} \rightarrow [-back]$$

$$(15) \begin{array}{c} \\ \\ p \ a \ p \ \acute{i} \ r \ n \ a \ k \end{array} \Rightarrow \begin{array}{c} [-back] \\ | \\ p \ a \ p \ \acute{i} \ r \ n \ a \ k \end{array}$$

We can now also explain the difference between *hég* and *kés* (restricting ourself to the possibility that [+back] is the active feature): the latter has underlying floating [+back], triggering harmony, whereas the former does not. Notice (i) that this requires a special subtheory of phonology, viz. underspecification theory, (ii) that this gives us a certain level of abstraction, but Kiparsky's main argument against abstraction disappears: it is not the case that *any* back vowel may be postulated, only the back vowel that differs minimally from the non-low unrounded vowel in question.

4 Constraints

There still is a problem regarding the previous approach: it requires some doubling. We have a clean-up rule which turns segments that cannot be underlying into segments which can. We have to postulate that there are no back unrounded high vowels in Hungarian and at the same time we need a rule system which first makes such vowels and then lets them disappear again. (This is very close to what is sometimes called the Duke-of-York gambit: "The grand old Duke of York he had ten thousand men / He marched them up to the top of the hill / And he marched them down again.")

Duplications like this have been part of the reason why constraint-based theories such as Optimality Theory have enjoyed such a popularity in phonology. These theories have also set our view of many things upside down. Look

for example at the outcome of the derivation in (13); what we see here, is an instance of harmony which has applied non-locally: it has skipped a vowel. But implicit in our discussion of the autosegmental and metrical alternatives to the representation of harmony was that skipping is not allowed. In the rule-based approach, our answer to this was that at the level at which harmony applied, there was no skipping. In an OT approach we could relativize the constraint involved (let us call it NOSKIPPING). We could say it is ranked below a constraint NOUNROUNDEDBACKVOWELS (say). By ranking the constraints in the inverse order, we get the Turkish system.

We also have to return to the issue of why lexically associated vowels do not spread. This is a topic which seems to be neither purely derivational in the sense of rules nor representational. The issue seems to be the following:

- (16) An association line may not be present in the output if it could have been potentially present in the input.

The lesson of this is that a good theory of vowel harmony needs both a good view of representations and a good view of derivations.

Bibliography

- Kenstowicz, Michael (1994). *Phonology in Generative Grammar*. Blackwell.
- Kiparsky, Paul (1982). 'How Abstract is Phonology?' In: *Explanation in Phonology*, no. 4 in *Publications in Language Sciences*, pp. 119–163. Foris.
- Lightner, Theodore (1965). 'On the Description of Vowel and Consonant Harmony'. *Word*, 21: 224–250.