The emergence of markedness in phonology

Paul Boersma
Groningen, December 8, 2006
The grammar model

The task of the listener: comprehension

The task of the speaker: production

(meaning)

| underlying form |

/surface form/

[auditory form]

[articulatory form]

(Boersma 2005, Apoussidou 2006)
The point: markedness can go

In this talk I will first illustrate that phonologists have traditionally invoked the concept of ‘markedness’ to explain a variety of phenomena in phonology, such as differences in frequency of occurrence and differences in degrees of phonological activity.

I will then defend the standpoint that markedness is not the underlying explanatory concept. Instead, differences in degrees of phonological activity can be shown to follow from differences in frequency of occurrence.
The emergentist explanation

I will show (with computer simulations of virtual learners) that if speakers use the same parallel multi-level Optimality-Theoretic grammar as listeners, the correlation between degree of phonological activity and frequency of occurrence automatically results from a simple learning algorithm in acquisition.
Example 1: the feature [±round]

Phonologists use the feature [±round] to distinguish between vowels made with lip rounding and vowels made without.

The discrete phonological feature value [+round] applies to vowels articulated with various degrees of lip rounding:

- [u] as in Dutch [buk] ‘book’
- [o] as in French [o] ‘water’
- [ɔ] as in German [ɡɔt] ‘God’
- [y] as in Turkish [ytʃ] ‘three’
- [ø] as in French [pø] ‘little’
- [œ] as in Limburgian [yœːt] ‘gutter’
- [ʊ] as in English [fʊːd] ‘food’
The discrete phonological feature value [–round] applies to vowels made with various degrees of lip spreading:

[i] as in Spanish [si] ‘yes’
[e] as in Italian [ke] ‘what’
[a] as in Russian [da] ‘yes’
[ɑ] as in English [ɑːsk] ‘ask’
[ɨ] as in Russian [sin] ‘son’
[ə] as in Chinese [móŋ] ‘door’
[ẽ] as in Portuguese [mẽnˈẽ] ‘morning’
[ɯ] as in Japanese [dzuːr:] ‘ten’

The two feature values [+round] and [–round] differ in two respects: frequency and strength.
Frequency differences for [±round]

In nearly all languages, the feature value [−round] is more common than the feature value [+round].

For instance: in an English text, 65% of the vowels are [−round], 35% are [+round].

(A possible cause: lip rounding causes all resonance frequencies to go down, so that sounds with lip rounding can be distinguished from each other a bit less well that sounds without lip rounding.)

But this talk does not try to explain this frequency difference.
Strength differences for \([\pm \text{round}]\)

\([+\text{round}]\) is phonologically strong, i.e. it can change \([-\text{round}]\):

\[\text{f}\varepsilon\text{l} + u \rightarrow \text{f}\o\text{elu} \; \text{‘much’} \; (\text{Saxon/Limburgian, 1000 A.D.})\]
\[([u] \text{ is both } [+\text{round}] \text{ and } [+\text{back}])\]
\[(\text{the } [+\text{round}] \text{ of } [u] \text{ spreads, although its } [+\text{back}] \text{ does not})\]

\([-\text{round}]\) is phonologically weak, e.g. it cannot change \([+\text{round}]\):

\[\text{fo:t} + i \rightarrow \text{f}\o\text{ti} \; \text{‘feet’} \; (\text{English/Saxon/Limburgian, 500 A.D.})\]
\[([i] \text{ is both } [-\text{round}] \text{ and } [-\text{back}])\]
\[(\text{the } [-\text{back}] \text{ of } [i] \text{ spreads, but its } [-\text{round}] \text{ does not})\]

Observation: the less common feature value (\([+\text{round}]\)) is the phonologically stronger one (i.e. it changes \([-\text{round}]\), and it is itself resistant to change).
Explanation by most phonologists:

1. “ [+round] is the marked feature value, and [−round] is the unmarked feature value.”

2. Universally, ‘marked’ feature values are less common than unmarked feature values.

3. Universally, ‘marked’ feature values are also stronger than unmarked feature values.

In other words, these phonologists invoke a sort of ‘aether’ as a universal cause of several phenomena: ‘markedness’ causes frequency differences, and ‘markedness’ causes strength differences.
Example 2: place of articulation

Phonologists use the feature [place] to distinguish between consonants made with various articulators (lips, tongue tip, tongue body).


Again, the three feature values [labial], [coronal], and [dorsal] differ in two respects: frequency and strength.

**Frequency differences for place**

In nearly all languages, the feature value [coronal] is much more common than the feature values [labial] and [dorsal], at least at the end of a word.

For instance: in an English text, word-final [n,t,d] are three times as common as word-final [m,p,b], and four times as common as word-final [ŋ,k,g].
Strength differences for place

[labial] and [dorsal] are phonologically stronger than [coronal]: they can change [coronal] to themselves, but they cannot be changed themselves. See the following examples from Dutch:

\[
\begin{align*}
\text{tn} + \text{pækən} & \rightarrow \text{impækən} \quad \text{(coronal changed by labial)} \\
\text{tn} + \text{kɛikən} & \rightarrow \text{ŋkɛikən} \quad \text{(coronal changed by dorsal)} \\
\text{ɔm} + \text{tɾɛkən} & \rightarrow \text{not ɔntrɛkən} \quad \text{(labial unchanged)} \\
\text{ɔm} + \text{kɛikən} & \rightarrow \text{not ŋkɛikən} \quad \text{(labial unchanged)} \\
\text{læŋ} + \text{parkeɾən} & \rightarrow \text{not lɑmparkeɾən} \quad \text{(dorsal unchanged)} \\
\text{dɪŋ} + \text{dun} & \rightarrow \text{not dɪndun} \quad \text{(dorsal unchanged)}
\end{align*}
\]

Observation: the most common feature value ([coronal]) is the weakest.
Explanation by most phonologists:

1. “[labial] and [dorsal] are the *marked* feature values, and [coronal] is the *unmarked* feature value.”

2. Universally, ‘marked’ feature values are less common than unmarked feature values.

3. Universally, ‘marked’ feature values are also stronger (more changing and less changeable) than unmarked feature values.

In other words, these phonologists invoke a sort of ‘aether’ as a universal cause of several phenomena: ‘markedness’ causes frequency differences, and ‘markedness’ causes strength differences.
I maintain that the concept of ‘markedness’ does not explain anything.

I claim: markedness (as the aether) does not exist.

I claim instead that frequency differences cause strength differences!
Synchronic functionalist explanation?

best perception is /m/  |  best perception is /n/

A perceived /m/ less reliably reflects an intended /m/ (light grey area divided by area left of criterion) than a perceived /n/ reflects an intended /n/ (dark grey divided by right of criterion).
Synchronic functionalist explanation?

The speaker is a listener as well, and ‘knows’ that the listener regards perceived /n/’s as reliable, and perceived /m/’s as unreliable.

The speaker can get away with producing an ‘underlying’ /n/ (as in Dutch /ɪn/) as an [m] (as in [ɪmpəkən]). The listener will probably perceive this sound as /m/ but regard it as unreliable, so that she can easily find the word /ɪn/ in her mental lexicon. By contrast, the speaker has to make sure not to produce an underlying /m/ (as in /ɔm/) as an [n] (as in *[ɔntɾɛkən]), because the listener would probably perceive this sound as /n/ and regard it as reliable, so that she will have trouble finding the word /ɔm/ in her mental lexicon (Boersma 1998, Hume 2004).
The emergentist explanation

The synchronic functionalist standpoint is too teleological (goal-oriented) to be real. Real brains do not have an omniscient supervisor, i.e. there is no ghost in the machine. What real brains do have is bidirectionality, which means that the same neural connections, with the same connection weights, are used for perception and for behaviour.

Less frequent sounds are stronger than frequent sounds, and the cause is that the speaker is also a listener: she uses in production the same grammar (connections) as she has established in the acquisition of comprehension.
The grammar model again

The task of the listener: comprehension

The task of the speaker: production

'meaning'

|underlying form|

/surface form/

[auditory form]

[articulatory form]
Acquiring comprehension

The first task of the learner is to acquire a good mapping from auditory form to underlying form. I assume for simplicity that the learner is provided with pairs of auditory and underlying forms (i.e. the lexicon has been built up correctly), and that she has to *invent* the phonological surface form.
Parallel comprehension in Optimality Theory

I assume (for the time being) that all processing is handled by Optimality Theory (Prince & Smolensky 1993), namely that:
(1) the relation between surface and underlying form is handled by faithfulness constraints (McCarthy & Prince 1995);
(2) the relation between auditory form and surface form is handled by cue constraints (Boersma 2005, Escudero 2005);
(3) that the two comprehension mappings occur in parallel.
Modelling the perception process

With ‘perception’ I mean a simplified version of parallel comprehension, namely one that does not involve the lexicon. Thus, perception is the mapping from an auditory form to the phonological surface form:

\[
\text{[auditory form]} \rightarrow \text{cue constraints} \rightarrow \text{surface form}
\]

(some psycholinguists regard this process as the first step of sequential comprehension, and therefore call it ‘prelexical perception’)

– 21 –
How cue constraints work in perception

Example: in English, there are at least two auditory cues to the voicing or voicelessness of a final obstruent: the presence or absence of periodicity (as in most languages), and the lengthening or shortening of the preceding vowel (this is specific to English). We can translate this into four cue constraints (“*” stands for “don’t have this”):

*+/voi/[-periodicity]

*–/voi/+[periodicity]

*/obs, +voi/[-lengthened vowel]

*/obs, –voi/+[lengthened vowel]
Two agreeing perception tableaus
Most often, the relevant cues agree, so that perception works well:

<table>
<thead>
<tr>
<th>[niːd]</th>
<th>*/obs, −voi/ [+lengthened vowel]</th>
<th>*/−voi/ [+periodicity]</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ˌnit./</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>/ˌnid./</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[niːt]</th>
<th>*/obs, +voi/ [−lengthened vowel]</th>
<th>*/+voi/ [−periodicity]</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ˌnit./</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ˌnid./</td>
<td>*!</td>
<td>*</td>
</tr>
</tbody>
</table>
A perception tableau with a conflict

But sometimes the cues disagree. Perception experiments have shown that in that case, the vowel lengthening constraint outranks the direct periodicity cue:

<table>
<thead>
<tr>
<th>[niːt]</th>
<th>*/obs,−voi/ [+lengthened vowel]</th>
<th>*/+voi/ [−periodicity]</th>
</tr>
</thead>
<tbody>
<tr>
<td>/.nit./</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>/./nid./</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

This is why in Optimality Theory the constraints are ranked from left to right in the tableau: a winning candidate does not have to satisfy all of the constraints.
Bidirectionality: phonetic production

The same cue constraints are used in production, i.e. in phonetic implementation.

\[ \text{/surface form/} \]

\[ \text{[auditory form]} \]

\[ \text{cue constraints} \]
Phonetic implementation with cue constraints only

The reuse of comprehension-based cue constraints in production is why speakers of English will try to have both cues right.

Here is an (incomplete) phonetic implementation tableau:

<table>
<thead>
<tr>
<th>/niːd./</th>
<th>*/obs,+voi/ [−lengthened vowel]</th>
<th>*/+voi/ [−periodicity]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[niːt]</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>[niːt]</td>
<td></td>
<td>*!</td>
</tr>
<tr>
<td>[niːd]</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>🔄 [niːd]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Parallel phonetic implementation

But phonetic implementation is not just about rendering cues. It is also about doing so efficiently, i.e. with the minimum expenditure of articulatory effort. Therefore, phonetic implementation is a parallel process that maps from a phonological surface form to a pair of auditory and articulatory form:

/surface form/

[cue constraints]

[auditory form]

[sensorimotor constraints]

[articulatory form]

[articulatory constraints]
Articulatory constraints

So we need articulatory constraints to evaluate the articulatory-phonetic form.

In the case at hand, we observe that it is especially difficult to pronounce periodicity in a final plosive. I express this simply as:

* [+periodicity, final plosive]

In a complete phonetic implementation tableau, this constraint must interact with cue constraints.
### Interaction of Articulatory and Cue Constraints

If the articulatory constraint outranks the lower-ranked cue constraint, speakers will implement only the most important cue:

<table>
<thead>
<tr>
<th>/ .nid./</th>
<th>*/obs,+voi/ [−lengthened vowel]</th>
<th>*[+periodicity, final plosive]</th>
<th>*/+voi/ [−periodicity]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[niːt]</td>
<td>*!</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>[niːt]</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>[niːd]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[niːd]</td>
<td></td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>

As we saw in the “conflicting perception” tableau, listeners still perceive this [niːt] as the intended / .nid./. This means speakers easily get away with saying [niːt]. And so they often say [niːt].
Case: nasal place assimilation

I will illustrate my point with the example of place assimilation of nasals, i.e. the case that an underlying |n| can be realized as /m/ but an underlying |m| cannot be realized as /n/.
The auditory language environment

The speaker realizes both an intended /n/ and an intended /m/ as a continuous place value along an auditory continuum, perhaps according to a Gaussian distribution. The distribution of the intended /n/ will be taller than that of the intended /m/:
Thus, there will be an infinite number of possible auditory place values. In my computer simulations, however, I simplifyingly assume that there are only four possible auditory forms for nasals, namely \([m], [M], [N]\), and \([n]\), where \([M]\) and \([N]\) are auditorily intermediate between \([m]\) and \([n]\).

**What are our cue constraints?**

They are the eight thinkable combinations of the four possible auditory forms and the two possible place feature values:

\[
*/n/[n] \quad */n/[N] \quad */n/[M] \quad */n/[m] \\
*/m/[n] \quad */m/[N] \quad */m/[M] \quad */m/[m]
\]
What are our faithfulness constraints?

The underlying form and the phonological surface form are made up of the same kind of material: discrete phonological elements such as segments, features, syllables, and feet. Faithfulness constraints aim at making the two forms identical:

|underlying form| faithfulness constraints

/surface form/

cue constraints

[auditory form]

For the case of place assimilation faithfulness constraints like the following are relevant: *|m|/n/ and *|n|/m/
Acquiring a constraint ranking

As said before, I assume that the child is given pairs of underlying form and auditory form. A possible input pair is therefore $|\text{an+pa}|[\text{aMpa}]$. For instance, the following tableau may occur at some point during acquisition:

| $|\text{an+pa}| [\text{aMpa}]$ | $*|n|/m|$ | $*/n/|M|$ | LAZY | $*/m/|M|$ |
|-----------------------------|---------|---------|-------|--------|
| $\sqrt{\text{an+pa}/.an.pa./[aMpa]}$ | | * | * |
| $|\text{an+pa}|/.am.pa./[\text{aMpa}]$ | *! | | | |

The form marked “√” is the one that is the most harmonic of all triplets that have both $|\text{an+pa}|$ and $[\text{aMpa}]$. The child will consider this the “correct” form.
Virtual production

After having established a “correct” triplet, the child computes the triplet that she herself would have produced, given the same underlying form |an+pa|.

| |an+pa| | *|n|/m/ | *|n/ [M] | LAZY | *|m/ [M] |
|---|---|---|---|---|---|---|
| √ |an+pa|/.an.pa./[aMpa] | | ! | * | |
| |an+pa|/.am.pa./[aMpa] | * | ! | * | * |
| ⚫ |an+pa|/.an.pa./[anpa] | | | *** | 

The “correct” triplet is also included in this tableau. The form marked “⚫” (the winner) is different from the form marked “√”, so that learning will occur.
The complete learning tableau

<table>
<thead>
<tr>
<th>an+pa</th>
<th>aMpa</th>
<th>*n/</th>
<th>*/n/</th>
<th>LAZY</th>
<th>*/m/</th>
</tr>
</thead>
<tbody>
<tr>
<td>√ an+pa/.an.pa./[aMpa]</td>
<td>*</td>
<td>*→</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>an+pa/.am.pa./[aMpa]</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>❸ an+pa/.an.pa./[anpa]</td>
<td></td>
<td>←***</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All the constraints that prefer the form marked “✸” will move by a small step down the constraint hierarchy (“→”), and all constraints that prefer the “correct” form (“✓”) will move up (“←”). This will make it more likely that in a future occurrence of |an+pa|[aMpa] the “✸” and “✓” forms will agree.
Computer simulation

A virtual child grows up in a language environment where coronals are twice as frequent as labials, and where people do not assimilate their nasals at all. However, the transmission channel introduces some variation (transmission noise), leading to the following underlying/auditory input distribution:

<table>
<thead>
<tr>
<th>form</th>
<th>frequency</th>
<th>form</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>an+pa</td>
<td>anpa]</td>
<td>72</td>
<td>am+ta</td>
</tr>
<tr>
<td>an+pa</td>
<td>aNpa]</td>
<td>62</td>
<td>am+ta</td>
</tr>
<tr>
<td>an+pa</td>
<td>aMpa]</td>
<td>44</td>
<td>am+ta</td>
</tr>
<tr>
<td>an+pa</td>
<td>ampa]</td>
<td>22</td>
<td>am+ta</td>
</tr>
</tbody>
</table>

Note: auditory assimilation is equally probable for coronals and labials, so the only coronal-labial difference is in the frequency.
The simulated learner (virtual baby) starts with all constraints ranked at the same height, and is subsequently fed a million input pairs randomly drawn from the input distribution given earlier. She learns with the learning algorithm given earlier.
Result: a frequency-based ranking

The virtual child ends up ranking her **faithfulness constraints**
directly be frequency of occurrence of the underlying feature value:

\[ \frac{|m|}{n} \text{ above } \frac{|n|}{m} \]
Bidirectional consequence

The child will use in production the same faithfulness constraints, with the ranking that she has learned in comprehension. They can interact with articulatory constraints, because phonological and phonetic production are handled in parallel:

faithfulness constraints

cue constraints

sensorimotor constraints

articulatory constraints

| underlying form |

/surface form/

[auditory form]

[articulatory form]
Bidirectional consequence

Given the frequency-based ranking $*|m|/n/ \gg *|n|/m/$, the learner has become more likely to assimilate a coronal nasal than to assimilate a labial nasal:

| |an+pa| | *|m|/n/ | LAZY | *|n|/m/ |
|---|---|---|---|---|---|
| /.an.pa./[anpa] | | *! | |
| /am.pa./[ampa] | * |

| |am+ta| | *|m|/n/ | LAZY | *|n|/m/ |
|---|---|---|---|---|---|
| /am.ta./[amta] | * | |
| /an.ta./[anta] | *! | |
Summary

Even if her parents do not have a preference for assimilating coronals over labials, the child will still develop such a preference as an automatic result of the frequency difference and the learning algorithm.

The cause of this is simply that the child uses the same constraint ranking in comprehension and production.

The concept of “markedness” is superfluous for explaining the greater phonological strength of $|m|$ as compared to $|n|$.
Prediction:

this theory predicts that in a language in which word-final dorsals happen to be as common as word-final coronals, the dorsals undergo change just as the coronals do.

In Austronesian languages, dorsals are indeed this common. Many words end in [ŋ], [ŋ], or [k]. And indeed, the dorsal [ŋ] changes easily by a labial or coronal:

**Indonesian:**

- məŋ + bɛri → membɛri
- məŋ + dapaŋ → mendapaŋ

**Tagalog:**

- pəŋ + butas → pembutas
- pəŋ + damo → pandamo
Cue reliability

Such computer simulations do more. Plosives have less transmission noise than nasals, leading to the following underlying/auditory input distribution:

<table>
<thead>
<tr>
<th>form</th>
<th>frequency</th>
<th>form</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>an+pa[anpa]</td>
<td>72</td>
<td>at+ma[atma]</td>
<td>114</td>
</tr>
<tr>
<td>an+pa[aNpa]</td>
<td>62</td>
<td>at+ma[aTma]</td>
<td>68</td>
</tr>
<tr>
<td>an+pa[aMpa]</td>
<td>44</td>
<td>at+ma[aPma]</td>
<td>16</td>
</tr>
<tr>
<td>an+pa[ampa]</td>
<td>22</td>
<td>at+ma[apma]</td>
<td>2</td>
</tr>
</tbody>
</table>

The simulation leads to a ranking by cue reliability:

*|t|/p/ above *|n|/m/
Conclusion

Some phenomena in phonology emerge as the result of an automatic acquisition bias.

If we can identify emergent phenomena, phonological theory will need less innatism and less synchronic functionalism.

Less innatism, because we need no innate ranking of \(^*|m|/n/\) over \(^*|n|/m/\) (‘coronal underspecification’) or \(^*|t|/p/\) over \(^*|n|/m/\) (Beckman 1998, Steriade 1995?).

Less synchronic functionalism, because we need not assume that the speaker explicitly takes into account the biases of the listener, neither for frequency (Boersma 1998, Hume 2004) nor for cue reliability (Steriade 1995?/2001, Boersma 1998).
The grammar model again

The task of the listener:
comprehension

The task of the speaker:
production

'meaning'

|underlying form|

/surface form/

[auditory form]

[articulatory form]