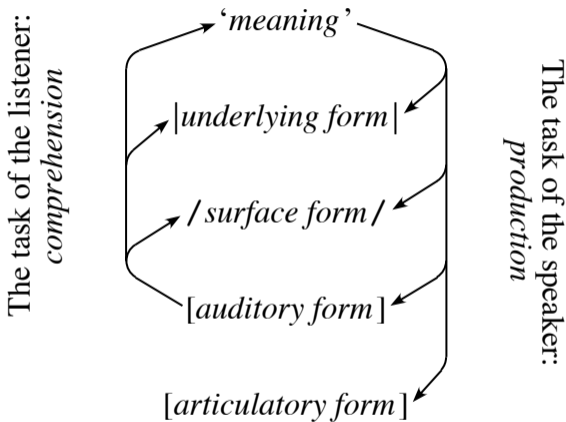# Parallel bidirectional phonology and phonetics

Paul Boersma
Belo Horizonte, November 29, 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 [gɔt] 'God'
[y] as in Turkish [ytʃ] 'three'
[ø] as in French [pø] 'little'
[œ] as in Limburgian [ɣœːt] 'gutter'
[ʉ] as in English [fʉːd] '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'
[ɛ] as in Swedish [ˇɛːta] 'eat'
[a] as in Russian [da] 'yes'
[ɑ] as in English [ɑːsk] 'ask'
[ɨ] as in Russian [sɨn] 'son'
[ə] as in Chinese [mə́n] 'door'
[ẽ] as in Portuguese [mẽɲˈẽ] 'morning'
[ɯ] as in Japanese [dʑɯː] '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 [±round]

[+round] is phonologically strong, i.e. it can change [−round]:

   fɛl + u → fœlu 'much' (Saxon/Limburgian, 1000 A.D.)
     ([u] is both [+round] and [+back])
     (the [+round] of [u] spreads, although its [+back] does not)

[−round] is phonologically weak, e.g. it cannot change [+round]:

   foːt + i → føːti 'feet' (English/Saxon/Limburgian, 500 A.D.)
     ([i] is both [−round] and [−back])
     (the [−back] of [i] spreads, but its [−round] does not)

Observation: the less common feature value ([+round]) is the phonologically stronger one (i.e. it changes [−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 'ether' 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).

The feature value [labial] applies to consonants made with the lips: [m] as in [hæm] 'ham', [p] as in [suːp] 'soup'.

The feature value [coronal] applies to consonants made with the tongue tip: [n] as in [hɛn] 'hen', [t] as in [wɒt] 'what'.

The feature value [dorsal] applies to consonants made with the tongue body: [ŋ] as in [ɹɪŋ] 'ring', [k] as in [hɛk] 'heck'.

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:

ɪn + pɑkən → ɪmpɑkən   (coronal changed by labial)
ɪn + kɛikən → ɪŋkɛikən   (coronal changed by dorsal)
ɔm + trɛkən → *not* ɔntrɛkən   (labial unchanged)
ɔm + kɛikən → *not* ɔŋkɛikən   (labial unchanged)
lɑŋ + pɑrkeːrən → *not* lɑmpɑrkeːrən   (dorsal unchanged)
dɪŋ + dun → *not* dɪndun   (dorsal unchanged)

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!

Proposed ultimate cause: human brains have *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*

'*meaning*'

|*underlying form*|

/ *surface form* /

[*auditory form*]

[*articulatory form*]
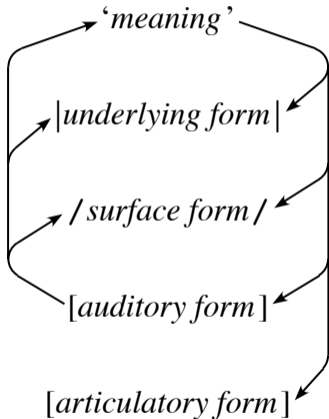
The task of the speaker: *production*

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.

|*underlying form*|

/ *surface form* /

[*auditory form*]
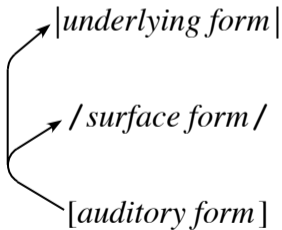
Parallel comprehension in Optimality Theory

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.
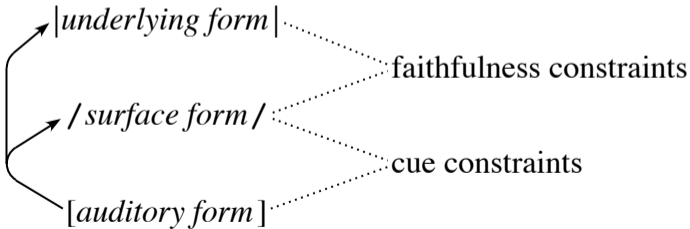
Modelling the perception process

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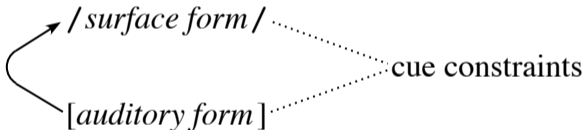
(some psycholinguists regard this process as the *first step* of sequential comprehension, and therefore call it 'prelexical perception')

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:

[niːd]	*/obs, −voi/ [+lengthened vowel]	*/−voi/ [+periodicity]
/.nit./	*!	*
☞ /.nid./		

| [niːt]	*/obs, +voi/ [−lengthened vowel]	*/+voi/ [−periodicity]
☞ /.nit./		
/.nid./	*!	*

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:

[niːt]	*/obs,–voi/ [+lengthened vowel]	*/+voi/ [–periodicity]
/.nit./	*!	
☞ /.nid./		*

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.
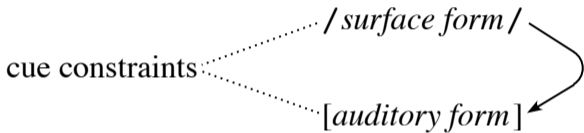
## Phonetic implementation with cue constraints only

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:

/.nɪd./	*/obs,+voi/ [−lengthened vowel]	*/+voi/ [−periodicity]
[niːt]	*!	*
[niːːt]		*!
[niːd]	*!	
☞ [niːːd]		

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:
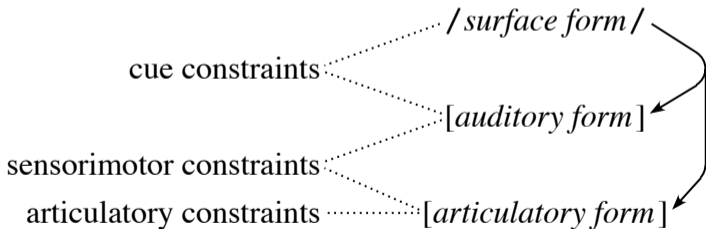
Articulatory constraints

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 implemantation 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:

| /.nid./	*/obs,+voi/ [−lengthened vowel]	*[+periodicity, final plosive]	*/+voi/ [−periodicity]
[niːt]	*!		*
☞ [niːːt]			*
[niːd]	*!		
[niːːd]		*!	

As we saw in the "conflicting perception" tableau, listeners still perceive this [niːːt] as the intended /.nid./. This means speakers easily get away with saying [niːːt]. And so they often say [niːːt].

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/:
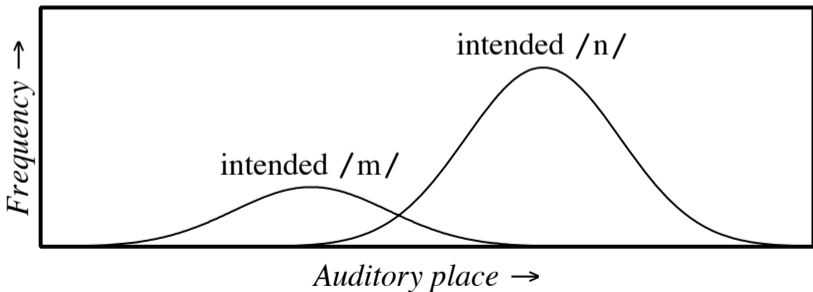
Thus, there will be an infinite number of possible auditory place values. In my computer simulations, however, I simplifyingly assume that there are only four possible auditory forms for nasals, namely [m], [M], [N], and [n], where [M] and [N] are auditorily intermediate between [m] and [n].

What are our cue constraints?

They are the eight thinkable combinations of the four possible auditory forms and the two possible place feature values:
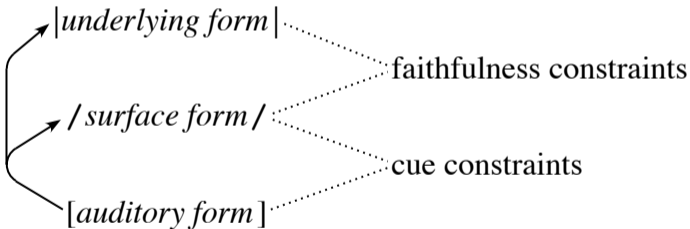
    */n/[n]   */n/[N]   */n/[M]   */n/[m]

    */m/[n]   */m/[N]   */m/[M]   */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:

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 |an+pa|[aMpa]. For instance, the following tableau may occur at some point during acquisition:

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

The form marked "√" is the one that is the most harmonic of all triplets that have both |an+pa| and [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

|an+pa| [aMpa]	*|n| /m/	*/n/ [M]	LAZY	*/m/ [M]
√ |an+pa|/.an.pa./[aMpa]		*→	*	
|an+pa|/.am.pa./[aMpa]	*		*	*
☞ |an+pa|/.an.pa./[anpa]			←***	

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:

form	*frequency*	*form*	*frequency*
|an+pa|[anpa]	72	|am+ta|[amta]	36
|an+pa|[aNpa]	62	|am+ta|[aMta]	31
|an+pa|[aMpa]	44	|am+ta|[aNta]	22
|an+pa|[ampa]	22	|am+ta|[anta]	11

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 listener ranks the **cue constraints** in such a way that she always chooses the category that was most likely to have been intended by the speaker. This minimizes confusion probability.
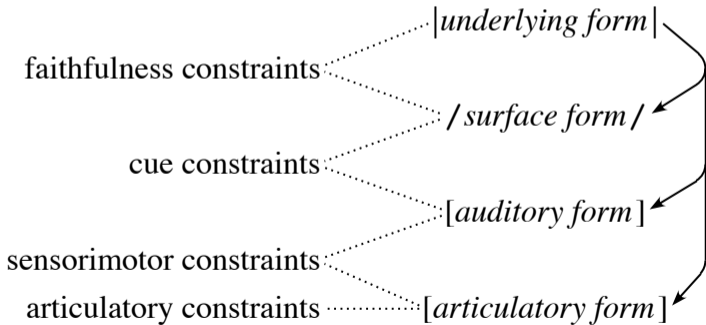

Auditory place →
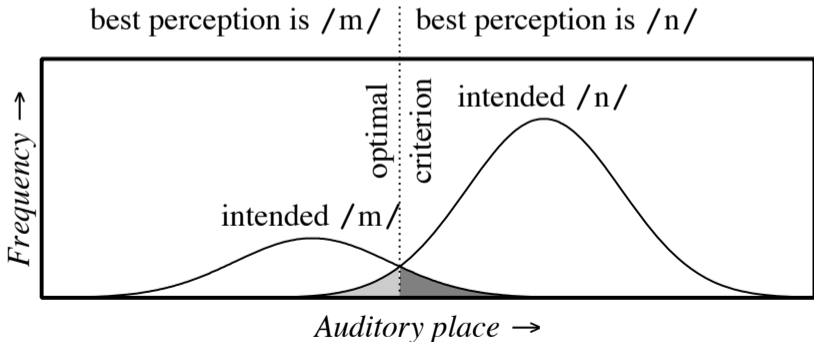
The virtual child ends up ranking her **faithfulness constraints** directly be frequency of occurrence of the underlying feature value:

$$*|m|/n/ \quad \text{above} \quad *|n|/m/$$

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

|*underlying form*|

/ *surface form* /

cue constraints

[*auditory form*]

sensorimotor constraints

articulatory constraints

[*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], [g], or [k]. And indeed, the dorsal [ŋ] changes easily by a labial or coronal:

Indonesian: məŋ + bəri → məmbəri
                məŋ + dapat → məndapat

Tagalog: pɛŋ + butɛs → pɛmbutɛs
                pɛŋ + dɛmo → pɛndɛmo

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/, neither an alternative account in terms of innate markedness constraints.
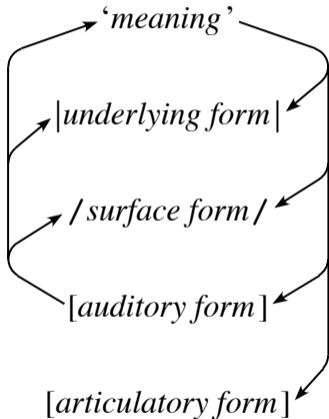
Less synchronic functionalism, because we need not assume that the speaker explicitly takes into account the biases of the listener.

References

Apoussidou, Diana (2006). On-line learning of underlying forms. *Rutgers Optimality Archive* **835**.

Boersma, Paul (2005). Some listener-oriented accounts of hache aspiré in French. *Rutgers Optimality Archive* **730**. Revised version to appear in *Lingua*.

Escudero, Paola (2005). *The attainment of optimal perception in second-language acquisition*. PhD thesis, Utrecht University.

McCarthy, John, and Alan Prince (1995). Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory.* University of Massachusetts Occasional Papers **18**. Amherst, Mass.: Graduate Linguistic Student Association. pp. 249–384. [*Rutgers Optimality Archive* **60**]

Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Technical Report TR-**2**, Rutgers University Center for Cognitive Science. [published in 2004]

The grammar model again

'*meaning*'

|*underlying form*|

/ *surface form* /

[*auditory form*]

[*articulatory form*]

The task of the listener:
comprehension

The task of the speaker:
production