

The acquisition and evolution of faithfulness rankings

Paul Boersma, *University of Amsterdam, paul.boersma@uva.nl*

Manchester Phonology Meeting, 27 May 2006

1. What this paper is going to show

This paper will derive observed universal rankings of faithfulness constraints from biases in acquisition that result from (1) frequency differences in the input and (2) imperfections in the transmission channel:

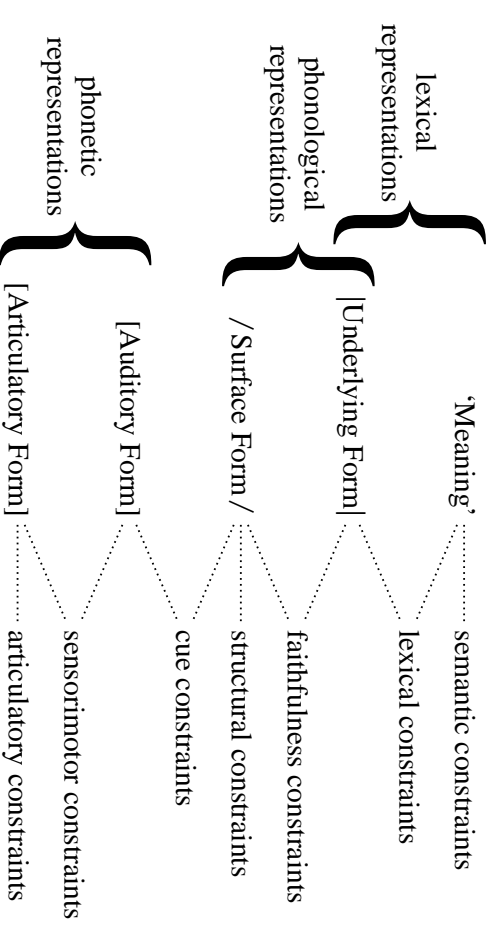
1.1. The computer simulations of the present paper show that in **acquisition**, the child's constraint rankings will *more or less* end up where they yield a grammar that generates the parents' language. But the ranking will diverge a bit, as a result of the above-mentioned frequency differences and imperfections.

1.2. Even if the parents' language has all faithfulness constraints ranked at the same height, their children will gradually rank them according to one of the **universal rankings** that have been proposed in the literature (e.g. *licensing by cue*, *positional faithfulness*, *probabilistic faithfulness*, *markedness as specification strength*).

1.3. All these rankings are therefore caused **automatically** by imperfections in the transmission channel in combination with a simple learning algorithm. None of the causes proposed before (all of which were based on the assumption that speakers have some sort of explicit or implicit linguistic or extralinguistic knowledge) are needed.

2. The required minimal bidirectional grammar model

2.1. These five **representations** are based on Boersma (1998, 2005), but feature an additional separation within the lexicon between the *signifié* and the *signifiant* (Saussure 1916), the two sides of the form-meaning pair (see also Diana Apoussidou's talk, this conference):



2.2. Seven constraint types: where do they come from?

The **structural** and **faithfulness** constraints are based on Prince & Smolensky (1993) and McCarthy & Prince (1995), the **articulatory** constraints on Jun (1995), Kirchner (1998), and Boersma (1998).

The **cue** constraints (the interface between phonology and phonetics) are based on Boersma (1998) and Escudero & Boersma (2003). The term was coined by Boersma (2005) and Escudero (2005).

The **lexical** and **semantic** constraints are based on Boersma (2001), Escudero (2005), and Apoussidou (this conference), who use them for choosing between underlying [rat] and [rad], [fʃekal] and [fʃikal], or [ðalas] and [ðalasl]. OT semanticists use similar form-meaning constraints for choosing between *kill* and *cause to die* (Blutner 2000), or between *him* and *himself* (Wilson 2001).

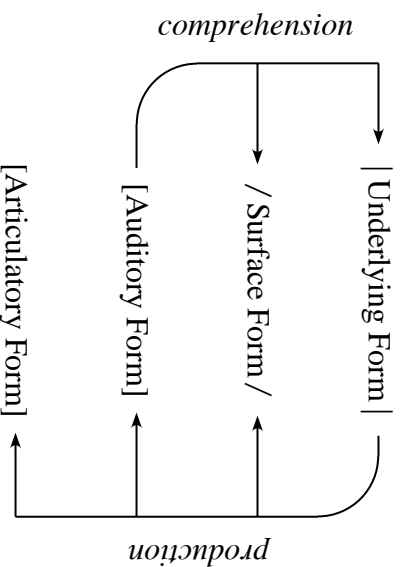
2.3. Simplification, just for this talk

For purposes of simplicity (but see the Discussion section), I will collapse Meaning and Underlying Form into one, basically assuming that there is a 1-to-1 relationship between the two (i.e. that the lexical constraints are ‘perfect’).

Also, I will usually collapse Auditory and Articulatory Form into one ‘phonetic form’, assuming that sensorimotor knowledge is ‘perfect’.

2.4. Bidirectional parallel processing

For this talk I simply assume that both comprehension and production are parallel processes (i.e. they do not consist of sequential modules):



2.5. Case: place assimilation

The forms that have to be evaluated in comprehension as well as in production are triplets of Underlying Form (e.g. |an+pa|, which includes a morpheme boundary), a discrete phonological Surface Form (e.g. /ampa/), and a phonetically detailed auditory/articulatory form (e.g. [aMpa], where [N] stands for a reduced coronal nasal). For instance, the triplet |an+pa| /ampa/ [ampa] shows full ‘phonological’ place assimilation, whereas the triplet |an+pa| /ampa/ [ampa] shows full ‘phonetic’ place assimilation.

3. Learning place assimilation from your parents

3.1. The primary language data for the beginning infant consists of auditory forms and a semantic/pragmatic context. For this talk, however, simulations begin in a later stage, namely when the child already has correct lexical form-meaning pairs. The primary language data consist, then, of **pairs of underlying form and auditory form**, e.g. |an+pa| [ampa].

3.2. Language environment for coronal nasal place assimilation.

For the simulations in chapter 3, I assume a Dutch or Catalan type of place assimilation: underlying |an+pa| is pronounced (by the parents) as [ampa] in 80% of the cases, and as [anpa] in 20% of the cases, while underlying |at+ma| is pronounced as [atma] in 99% of the cases, and as [apma] in 1% of the cases. Underlying forms with labial codas are pronounced fully labially. The full **distribution** of underlying–auditory pairs is then:

an+pa [ampa]	20	an+ta [anta]	100
an+pa [anpa]	80	an+ta [anta]	0
at+ma [atma]	99	ap+na [apna]	100
at+ma [apma]	1	ap+na [atna]	0

Question 1: will the simulated learners mimic these distributions?

3.3. The learning task is to find the intermediate forms, namely the **surface forms**. Like other surface structures (e.g. foot structure), these forms are inaudible and are not part of the primary language data; they have to be constructed by the learner. For the learning pair |an+pa||ampa|, the child probably constructs either /ampa/ or /anpa/.

Question 2: what will the simulated learners’ surface forms look like?

3.4. The learning algorithm is a simple *bidirectional optimization*, which is illustrated in the tableau below, where an illustrative example grammar (different from the ones used in my simulations) contains one faithfulness constraint, one unlikely structural constraint, four cue constraints, and three articulatory constraint.

The tableau illustrates the learning of the pair |an+pa| [ampa] and contains all candidates that contain either |an+pa| or [ampa] (or both). The numbers above the tableau show the continuous ‘disharmonies’ (ranking values + evaluation noise) of every constraint (I assume Stochastic OT; see Boersma 1997 and Boersma & Hayes 2001).

The candidate that the learner will regard as ‘**correct**’ (✓) is the one that is the most optimal of all the candidates that contain both |an+pa| and [ampa]; the ‘**forward**’ winner (☞) is the best of all the candidates that contain |an+pa|; and the ‘**backward**’ winner (☜) is the best of all the candidates that contain [ampa].

The tableau shows how the constraints will **move** if one or both of the winners differs from the ‘correct’ candidate. All constraints violated in an ‘incorrect’ winner will move up the continuous ranking scale, and all constraints violated in the ‘correct’ form will move down (by two steps in the tableau, because the form is compared to two other forms).

	an+pa ampa	*/p/[t]	*/t/[p]	IDENT	*/n/[m]	*/m/[n]	*[lip-tongue synchronization]	*[lip gesture]	*/mp/	*[tongue gesture]
✓	an+pa / anpa/ [ampa]				*→→			*→→		
☞	an+pa / anpa/ [ampa]						←*	←*		←*
	an+pa / ampa/ [ampa]			*			*	*	*	
	an+pa / ampa/ [ampa]			*	*		*	*	*	*
	an+pa / anpa/ [ampa]			*	*		*	*	*	*
☜	an+pa / ampa/ [ampa]							←*	←*	

3.5. Simplification number 5
 After the previous four simplifications (collapsing word-meaning pairs, collapsing auditory and articulatory forms, assuming parallel comprehension, and assuming the single learning algorithm of bidirectional optimization), there is a fifth simplification.

This simplification, not implemented in the tableau below, is that production does not neutralize, i.e. the underlying form can always be recovered from the auditory form (or from the surface form). For instance, there will exist an underlying form |an+pa|, but there will not exist an underlying form |an+pa| such as appears in the tableau below.

With this simplification, bidirectional optimization reduces to ‘forward’ optimization (no “☜” in the learning tableaux), i.e. robust perception with virtual production (Aoussidou & Boersma 2004). This is fully comparable to Tesar & Smolensky’s (1998, 2000) ‘robust interpretive parsing with constraint demotion’, except that the mappings from Surface Form to Auditory Form (Tesar & Smolensky’s Overt Form) and from Surface Form to Underlying Form are in the present paper no longer trivial but handled by the grammar, i.e. by the rankings of cue constraints and faithfulness constraints, respectively.

3.6. Simulation: a single constraint for place faithfulness

The tableau below shows simulation results for a learner with four cue constraints (the one on top means “a phonological element /p/ does not correspond to a phonetic [t]”), one articulatory constraint LAZY (violated by the articulatory distance between the coda and the onset, in some arbitrary units), and a single faithfulness constraint IDENTPLACE.

The learner is restricted to the four underlying forms shown. After 400,000 learning pairs drawn from the distribution in §3.2, this learner

ends up with the ranking values above the tableau (training scheme: 100,000 learning pairs each at plasticities of 1, 0.1, 0.01, and 0.001; the strength of the evaluation noise is always 2.0). The numbers at the right are output frequencies given the underlying form |an+pa| (or |at+ma|), computed by running it through the stochastic grammar 100,000 times.

Answers: with these constraints and candidates, learners replicate their parents’ distribution; and with a single IDENTPLACE, the surface form is always /ampa/, i.e., assimilation of [n] must be entirely phonetic.


	105.9	105.3	102.6	100.9	93.9	91.5	
	*[t] /p/	*[n] /m/	IDENTPLACE	*[p] /t/	LAZY	*[m] /n/	
☞ an+pa / anpa / [ampa]						*	an+pa :
an+pa / anpa / [ampa]					*!***		80%
an+pa / anpa / [ampa]					*!*		20%
an+pa / anpa / [ampa]		*!	*		***		0%
an+pa / anpa / [ampa]			*!		***	*	0%
am+ta / anta / [anta]							
am+ta / anta / [anta]						*!***	
am+ta / amta / [amta]							
am+ta / amta / [amta]		*!					at+ma :
at+ma / atma / [apma]				*!			0.7%
at+ma / atma / [atma]						*!***	99.2%
at+ma / apma / [apma]			*!				0.1%
at+ma / apma / [atma]		*!	*		***		0%
ap+na / atna / [apna]					*!		
ap+na / atna / [atna]							
ap+na / apna / [apna]						*!***	
ap+na / apna / [atna]							
ap+na / apna / [atna]		*!					

3.7. Simulation: more phonology by more granular faithfulness

To achieve more ‘phonological’ solutions, this tableau implements a possibility inspired by Jun (1995), namely to have two separate faithfulness constraints for the two values of the place feature, namely IDENT_{PLACE}([labial]) and IDENT_{PLACE}([coronal]), and to have two separate faithfulness constraints for the two different manners, namely IDENT_{PLACE}(nasal) and IDENT_{PLACE}(plosive).

Answers: with these constraints and candidates, the learners replicate their parents’ distribution. The 81% [ampa] forms are now distributed more or less evenly between ‘phonetic’ assimilation (candidate 1, /ampa/) and ‘phonological’ assimilation (candidate 3, /ampa/).

Note: the pointing finger at the candidate [an+pa]/[ampa]/[ampa] just means that this is the most harmonic triplet in the learner’s language.

	107.0	106.9	106.9	102.3	101.7	94.8	93.8	93.5	88.6
	*[n] /m/	*[t] /p/	IDPL (lab)	*[p] /t/	IDPL (plos)	LAZY	IDPL (nas)	*[m] /n/	IDPL (cor)
 an+pa / anpa/ [ampa]								*	
an+pa / anpa/ [ampa]						*!***			
an+pa / anpa/ [ampa]							*!		*
an+pa / anpa/ [ampa]									*
an+pa / anpa/ [ampa]	*!					***	*	*	*
an+pa / anpa/ [ampa]			*!			***	*	*	*
an+pa / anpa/ [ampa]									
an+pa / anpa/ [ampa]						*!***			
an+pa / anpa/ [ampa]									
an+pa / anpa/ [ampa]	*!								
an+pa / anpa/ [ampa]				*!					
an+pa / anpa/ [ampa]						*!***			*
an+pa / anpa/ [ampa]									
an+pa / anpa/ [ampa]		*!				***			*
an+pa / anpa/ [ampa]				*		***			
an+pa / anpa/ [ampa]			*!						
an+pa / anpa/ [ampa]						*!***			
an+pa / anpa/ [ampa]									
an+pa / anpa/ [ampa]		*!							

|an+pa| :

45%

19%

36%

0%

|at+ma| :

0.4%

98.9%

0.7%

0%

3.8. Simulation: variation between learners

This simulation follows Boersma (1998: 217) in having the faithfulness constraints “IDPL” (= IDENT_{PLACE}) specific to each of the four underlying segments. The cue constraints are still universal (no *[m]/m/) and discrete (no *[M]/m/). This means that the cue constraints and the faithfulness constraints have the exact same degree of granularity (four) and the exact same bias (towards “identity”).

Perhaps not surprisingly, *[m]/n/ and IDENT_{PLACE}(n) end up being ranked equally high. For the combination of underlying |an+pa| and phonetic [ampa], the surface form will therefore be /ampa/ half of the time (‘phonetic assimilation’) and /anpa/ half of the time (‘phonological assimilation’). **This simulated learner** appears to have the same trouble localizing assimilation as the linguistic community has as a whole.

	105.9	105.6	105.5	104.9	100.4	100.4	93.2	92.1	92.0
	IDPL (m)	*[t] /p/	*[n] /m/	IDPL (p)	IDPL (t)	*[p] /t/	LAZY	*[m] /n/	IDPL (n)
an+pa / anpa/ [ampa]								*!	
an+pa / anpa/ [anpa]							*!*		
☞ an+pa / ampa/ [ampa]			*!				***		*
an+pa / ampa/ [anpa]							***	*	
am+ta / anta/ [amta]		*!					***	*	
am+ta / anta/ [anta]									
am+ta / amta/ [amta]							*!*		
am+ta / amta/ [anta]			*!						
at+ma / atna/ [apma]						*!			
at+ma / atrna/ [atrna]							*!*		
at+ma / apma/ [apma]					*!				
at+ma / apma/ [atrna]		*!			*		***		
ap+na / atrna/ [apna]						*	***		
ap+na / atrna/ [atrna]				*!					
ap+na / apna/ [apna]								*!*	
ap+na / apna/ [atrna]		*!							

|an+pa| :

40%

20%

40%

0%

|at+ma| :

0.5%

99%

0.5%

0%

But some learners end up with a grammar that does ‘phonetic assimilation’) all of the time, at least for the nasals. The phonetic forms still match the input distribution, though. (for the plosives, it’s still half-half)

Differently from the previous learner, **this simulated learner** appears to have made the same choice as a part of the linguistic community.

	105.3	105.0	103.9	103.4	101.9	99.8	98.9	92.0	89.7
	IDPL (p)	*[n] /m/	*[t] /p/	IDPL (m)	IDPL (n)	*[p] /t/	IDPL (t)	LAZY	*[m] /n/
☞ an+pa / anpa / [anpa]									*
an+pa / anpa / [anpa]								*!***	
an+pa / ampa / [ampa]					*!				
an+pa / ampa / [ampa]			*!		*			***	
an+pa / anta / [anta]				*!				***	*
an+ta / anta / [anta]									
an+ta / amta / [amta]								*!***	
am+ta / anta / [anta]			*!						
at+ma / atma / [atma]						*!			
at+ma / atma / [atma]								*!***	
at+ma / apma / [apma]							*!		
at+ma / apma / [atma]			*!				*	***	
ap+na / atna / [apna]		*!				*		***	
ap+na / atna / [atna]									
ap+na / apna / [apna]								*!***	
ap+na / apna / [atna]			*!						

|an+pa| :

80%

20%

0%

0%

3.9. Simulation: auditory cues are arbitrary

If we regard auditory cues as arbitrarily related to phonological elements, then we have to include cue constraints like $*[m]/m/$, because the sound $[m]$ has no a priori relation with the phoneme $/m/$. In the simulation, something goes out of hand: LAZY shoots off beyond the bottom of the hierarchy. This is because $*[n]/n/$ will now outrule the ‘all-ampa’ candidate.

Answers. These learners do *not* match their parents’ distribution.

To allow the presence of arbitrary constraints (including seemingly contrary ones like $*[n]/n/$), we will need some sanity data such as $[in+ti]$ $[inti]$. (note: such sanity data could also come from the onset of the second syllable in each form here, but I assume that all the constraints mentioned here are specific to the coda, so that $/apna/[apna]$ does not violate $*[n]/n/$)

	106.4	104.5	104.4	103.2	101.5	101.4	98.6	98.5	97.4	94.8	51.6	51.6	-5558.6
	IDPL (p)	$*[p]/t/$	IDPL (t)	$*[t]/p/$	$*[n]/m/$	$*[n]/n/$	$*[m]/n/$	$*[m]/m/$	$*[t]/t/$	$*[p]/p/$	IDPL (n)	IDPL (m)	LAZY
$ an+pa / anpa / [ampa]$							$*!$						
$ an+pa / anpa / [anpa]$						$*!$							***
$ an+pa / ampa / [ampa]$								$*!$			*		***
$ an+pa / ampa / [anpa]$					$*!$						*		***
$ am+ta / anta / [anta]$							$*!$				*		***
$ am+ta / anta / [anta]$												*	***
$ am+ta / amta / [amta]$								$*!$					***
$ am+ta / amta / [anta]$					$*!$								***
$ at+ma / atma / [apma]$													***
$ at+ma / atma / [atma]$									$*!$				***
$ at+ma / apma / [apma]$						$*!$				*			***
$ at+ma / apma / [atma]$										*			***
$ ap+na / atna / [apna]$										*			***
$ ap+na / apna / [apna]$										*			***
$ ap+na / apna / [atna]$				$*!$									***

$|an+pa| :$

45%

5%

45%

5%

3.10. Simulation: more underlying forms, for more reliable learning

When an equal number of non-assimilating forms (exemplified by forms with the vowel /i/) is included (all faithful and with perfect cues), we again get an 80% assimilation for |n|, and it is again due to an equal ranking of *[m]/n/ and IDENTPLACE(|n|) just below LAZY. A 1% assimilation for |t| is again due to an equal ranking of *[p]/t/ and IDENTPLACE(|t|) 3 noise strengths above LAZY.

Conclusion: it's important to have sufficiently broad primary language data.

	122.6	122.6	122.1	121.7	118.6	118.5	111.1	110.4	109.9	101.5	85.6	73.8	65.6
	IDPL (p)	*[n] /m/	*[t] /p/	IDPL (m)	*[p] /t/	IDPL (t)	LAZY	*[m] /n/	IDPL (n)	*[n] /n/	*[t] /t/	*[p] /p/	*[m] /m/
an+pa / anpa / [ampa]								*!					
an+pa / anpa / [anpa]							*!***			*			
an+pa / ampa / [ampa]									*!				*
an+pa / ampa / [anpa]		*!					***		*				
am+ta / anta / [amta]				*!			***	*					
am+ta / anta / [anta]				*!						*			
am+ta / amta / [amta]							*!***						*
am+ta / amta / [anta]		*!											
at+ma / atma / [apma]					*!								
at+ma / atma / [atma]							*!***				*		
at+ma / apma / [apma]						*!						*	
at+ma / apma / [atma]			*!			*	***						
ap+na / atna / [apna]	*!				*		***						
ap+na / atna / [atna]	*!									*			
ap+na / apna / [apna]							*!***					*	
ap+na / apna / [atna]			*!										
☞ im+pi / impi / [impi]													*
im+pi / impi / [inpi]		*!					***						
im+pi / inpi / [impi]				*!				*					
im+pi / inpi / [inpi]				*!			***			*			
in+ti / inti / [imti]							*!***	*					
in+ti / inti / [inti]										*!			
in+ti / imti / [imti]							*!***		*				*
in+ti / imti / [inti]		*!							*				
ip+mi / ipmi / [ipmi]												*!	
ip+mi / ipmi / [itmi]			*!				***						
ip+mi / itmi / [ipmi]	*!				*								
ip+mi / itmi / [itmi]	*!						***				*		
it+ni / itni / [ipni]					*!		***						
it+ni / itni / [itni]										*!			
it+ni / ipni / [ipni]						*!	***					*	
it+ni / ipni / [itni]			*!			*							

3.11. Remaining problem: real auditory cue values are not unary, not binary, but multi-valued.