Emergent ranking of faithfulness explains markedness and licensing by cue[*]

Paul Boersma, University of Amsterdam, 10 March 2008

Abstract. This paper derives 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. Computer simulations 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 said frequency differences and imperfections. 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*, *markedness by faithfulness*). All these rankings are therefore caused automatically by 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.

The typology of place assimilation contains the following implicational universals (Mohanan 1993):

(1) *Typology of place assimilation*
  a. if a velar or labial assimilates, so does its coronal counterpart;
  b. if a plosive assimilates, so does its nasal counterpart.

Observation (1a) has been explained by MARKEDNESS: coronal is the unmarked place, therefore coronals are more likely than other places to obtain their place from a neighbour (Paradis & Prunet 1991). Observation (1b) has been explained by LICENSING BY CUE: speakers know that phonological place values are easier to distinguish in plosives than in nasals, so that they know that listeners will accept place changes more readily in nasals than in plosives (Jun 1995; Steriade 1995, 2001; Boersma 1998).

The present paper *derives* the typology in (1) from a simple learning procedure. As for (1a), I show that markedness relations derive automatically from frequency differences in acquisition, so that the concept of markedness does not have to be invoked as an independent explanatory linguistic device. As for (1b), I show that universal rankings of faithfulness constraints by auditory cue quality can come about automatically during acquisition, so that listeners need have no direct knowledge of degrees of auditory distinctivity. The results generalize to many more areas of phonology than just place assimilation.

The paper is organized as follows. Section 1 introduces the theoretical machinery that is required to make my explanations work. Section 2 provides a validation of the

---

[*] This paper was first presented at the Manchester Phonology Meeting on 27 May 2006, and benefited from audiences at SIGPhon (New York, 8 June 2006), the LOT Summerschool (Amsterdam, 12–16 June 2006), the Jadertina Summer School in Empirical and Computational Linguistics (Zadar, 18–22 September 2006), the English Department of the Universidade Federal de Santa Catarina (Florianópolis, 24 November 2006), the III Congresso Internacional de Fonética e Fonologia (Belo Horizonte, 29 November 2006), the Workshop on Computing and Phonology (Groningen, December 8, 2006), the Workshop on Experimental Approaches to OT (Ann Arbor, 19 May 2007), a lecture series at the Center for Advanced Study in Theoretical Linguistics (12–14 November 1007), the Colloquium Taal en Spraak (Nijmegen, 20 February 2008), and the Linguistics Department of the Massachusetts Institute of Technology (29 February 2008).

proposed learning algorithm. Section 3 performs computer simulations that show that even in a language environment where adults exhibit no differences in assimilation between coronals and other places, nor between plosives and nasals, their children will automatically end up being more likely to assimilate their coronals than their labials (if coronals are more frequent) and more likely to assimilate their nasals than their plosives (if plosives have better auditory cues). Section 4 tests some empirical predictions of the proposal, and shows that many phenomena formerly ascribed to markedness, positional faithfulness, or licensing by cue, can also just be regarded as resulting from biases in acquisition that are predictable from differences in frequency and audibility.

1 The minimally required bidirectional grammar model

1.1 Representations and their connections

The purpose of the present paper requires taking into account the three representations shown in Fig. 1, which are based on a more elaborate model by Boersma (1998, 2007a).

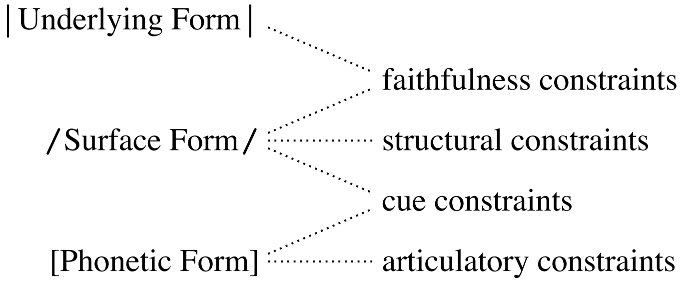

Figure 1
Representations and connections.

Two of the three representations in Fig. 1 are phonological. I understand them here in the traditional way. Thus, an Underlying Form is a sequence of discrete phonological structures associated with morphemes in the lexicon. For the place assimilation case, a relevant underlying form could be |in+pa|, a sequence of two lexical forms connected by the morpheme boundary '+'. A Surface Form is a discrete phonological surface structure consisting of segments, features, syllables, and feet. For the place assimilation case, relevant surface forms could be /.im.pa./ and /.in.pa./; in these simplified representations, the period '.' denotes a syllable boundary.

The third representation in Fig. 1 is phonetic. One might like to distinguish between an auditory and an articulatory representation. An auditory representation is a sound filtered by the listener's peripheral auditory system: it contains continuous data such as pitches, formants, silences, and noises. An articulatory representation is a set of muscle commands; it contains continuous data such as the activities of the muscles of the tongue, lips, velum, pharynx, larynx, and lungs. For the purposes of the present paper I can simplifyingly collapse these two representations into a single Phonetic Form, assuming that speakers have perfect sensorimotor knowledge, i.e. that they maintain an appropriate one-to-one relation between sound and articulation. For the place assimilation case, some relevant phonetic forms could be [impa], [iMpa], [iNpa], or [inpa]; in these shortcut representations, [M] represents an auditorily and articulatorily reduced labial nasal, and [N] a reduced coronal nasal.

To describe a production in terms of the three representations of Fig. 1, we have to specify a triplet of Underlying, Surface, and Phonetic Form. For instance, a fully faithful production with a coronal nasal underlyingly, at the surface, and in the phonetics, is represented by the triplet |in+pa| /.in.pa./ [inpa]; a case of phonological place assimilation, where the surface structure already has a phonologically labial nasal, can be represented by the triplet |in+pa| /.im.pa./ [impa]; a case of full 'merely phonetic' place assimilation, where the phonological surface structure has a coronal nasal that is phonetically pronounced by closing the lips without tongue constriction, can be represented by the triplet |in+pa| /.in.pa./ [impa]; and a case of partial phonetic place assimilation, where the nasal is pronounced as a reduced coronal, can be represented by the triplet |in+pa| /.in.pa./ [iNpa].

The three representations are evaluated by four types of constraints. Faithfulness constraints evaluate differences between Underlying and Surface Form (McCarthy & Prince 1995). Structural constraints evaluate phonological surface structures (Prince & Smolensky 1993). Articulatory constraints evaluate articulatory effort (Jun 1995, Steriade 1995, Kirchner 1998, Boersma 1998). Cue constraints form the interface between phonology and phonetics; they express the speaker-listener's knowledge of the relations between continuous auditory cues and discrete phonological surface elements (Escudero & Boersma 2003, 2004; Escudero 2005; Boersma 2006, 2007ab). I next illustrate the workings of these constraints.

1.2 Phonology and phonetics in parallel

One way to define a production procedure on the basis of Fig. 1 is to propose that the speaker, when given an Underlying Form to produce, first maps this Underlying Form to a Surface Form ("phonological production"), and then uses the resulting output (the surface form) as the input to a second process ("phonetic implementation"), in which the speaker maps this Surface Form to a Phonetic Form. In this serial view of phonological-phonetic production, place assimilation could take place in either of the two processes. "Phonological" place assimilation would take place if the relevant faithfulness constraint is outranked by a structural constraint, e.g. SpreadPlace (Padgett 1995), that dislikes adjacent non-identical place values, as in (2).

(2) *Phonological assimilation*

|in+pa|	SpreadPlace	IdentPlace(|n|)
/.in.pa./	*!	
☞ /.im.pa./		*

Tableau (2) is not different from most accounts of the interaction between structural and faithfulness constraints in the OT literature since McCarthy & Prince (1995). The faithfulness constraint, however, is relativized for the particular values of place and manner, i.e. there are separate faithfulness constraints for underlying |n|, |m|, |t|, and |p|. This granularity of faithfulness constraints is required to allow us to describe the typology in (1) in terms of constraint rankings. For instance, the ranking IdentPlace(|m|) >> SpreadPlace >> IdentPlace(|n|) describes a language in which |n| assimilates (phonologically) but |m| does not.

The second potential locus of assimilation is the phonetic implementation component. Even if the phonology keeps the coronal nasal intact, i.e. it maps |in+pa| to /.in.pa./, phonetic implementation is capable of performing assimilation, as in (3).

(3) *Phonetic assimilation*

/.in.pa./	*[TongueTip]	*/n/[m]
[inpa]	*!	
☞ [impa]		*

In almost complete analogy with the phonological process in (2), the assimilation in (3) is achieved by the articulatory constraint *[TongueTip] outranking the cue constraint */n/[m]. The articulatory constraint is violated in the first candidate only: the articulation [inpa] involves both a tongue-tip gesture (for [n]) and a lip gesture (for [p]), whereas the articulation [impa] involves only a lip gesture, which is shared by [m] and [p]. The cue constraint */n/[m] expresses the idea that the phonological structure /n/ (i.e. a combination of the features nasal and coronal) does not connect well to the sound [m] (i.e. loud periodicity, low second formant transitions, spectral zero); this is violated only in the second candidate.

An interesting possibility not representable in the serial production model of (2) and (3) is that the lower-level articulatory constraint *[TongueTip] forces phonological assimilation, i.e. the violation of the higher-level faithfulness constraint IdentPlace(|n|). Precisely such proposed cases of phonology-phonetics interaction, in which the “later” phonetics influences the “earlier” phonology, made Jun (1995), Steriade (1995) and Kirchner (1998) propose grammar models in which surface forms had phonological as well as phonetic aspects. As Boersma (1998) showed, however, the distinction between discrete phonological and continuous phonetic levels of representation can be maintained without sacrificing the influence of the phonetics on the phonology. While Boersma (1998) achieved this within a control loop model of phonology, it can also simply be achieved by regarding the phonological mapping (from Underlying to Surface Form) and the phonetic mapping (from Surface to Phonetic Form) as parallel and interactive processes (Boersma 2007ab). The relevant tableau is (4).

(4) *Phonological assimilation caused by the phonetics*

|in+pa|	*/n/[m]	*[TongueTip]	IdentPlace(|n|)
/.in.pa./ [inpa]		*!	
/.in.pa./ [impa]	*!		
☞ /.im.pa./ [impa]			*

Thus, if the “phonological” constraint IdentPlace(|n|) is outranked by the “phonetic” constraint *[TongueTip] and the “phonology-phonetics interface” constraint */n/[m], the phonetics will force the phonology to perform place assimilation.

One reason for positing parallel multi-level processing as in (4) is economy: if a grammar model requires articulatory constraints anyway (to model phonetic implementation), it is most economical to have them influence the phonology as in (4), because then the grammar model can dispense with (some) structural constraints.

However, the present paper makes no attempt to argue for or against the parallel model on such grounds or on empirical grounds; for general arguments for bottom-up influence of phonetics on phonology in production, the reader can consult Jun (1995), Flemming (1995), Steriade (1995, 2001), Kirchner (1998), and Boersma (1998), and for specific empirical arguments in favour of the parallel model, the reader can consult Boersma (2007a). What the present paper does aim at showing is that the parallel model exemplified in (4) explains phenomena formerly ascribed to markedness relations and to knowledge of auditory distinctivity; it does so only in combination with a learning procedure, which I describe next.

1.3 Learning to assimilate

The learning algorithm requires that production is a parallel process (i.e. it does not consist of sequential modules), as in Fig. 2. The requirements on comprehension may be less severe, although Fig. 2 depicts comprehension as a parallel process as well.

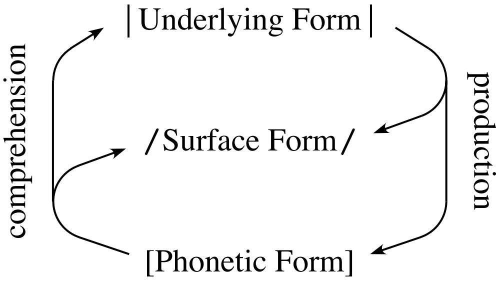

Figure 2
Parallel production and comprehension.

In general, PRIMARY LANGUAGE DATA for the beginning infant consist of auditory forms and a semantic-pragmatic context. For the present paper, however, simulations begin in a later stage, namely when the child already has correct lexical form-meaning pairs. That is, I assume (somewhat articificially) that the child has a complete lexicon, and that she knows the Underlying Form of every lexical item. For this simplified child, then, the primary language data consist of pairs of underlying form and phonetic form, e.g. |in+pa| [impa].

Imagine now a child who has faithfulness constraints, cue constraints, and articulatory constraints (and hence no structural constraints). Imagine further that at some point during acquisition, the child has a grammar that does not assimilate, as in (5).

(5) *Learning assimilation*

|in+pa| [impa]	*/n/[m]	IDENTPLACE(|n|)	*[TONGUETIP]
☞ /.in.pa./ [inpa]			←*
/.in.pa./ [impa]	*!		
√ /.im.pa./ [impa]		*!→	

In the absence of structural constraints, the only way to obtain assimilation is to have the articulatory constraint ranked above either the faithfulness constraint (so that one obtains phonological assimilation) or above the cue constraint (so that one obtains phonetic assimilation). In (5), therefore, non-assimilation is guaranteed by the fact that both the faithfulness constraint and the cue constraint are ranked above the articulatory constraint. Therefore, the winning candidate in the child's production, given the underlying form |in+pa|, is the surface-phonetic pair /.in.pa./ [inpa], as indicated by the pointing finger in (5).

Tableau (5) contains more information than just the underlying form: (5) is a LEARNING TABLEAU. Tableau (5) occurs when the child hears an adult producing the sound [impa]. The child's language environment, therefore, is one in which adults do assimilate. I assume that the child can induce that the speaker intended to produce the underlying form |in+pa|, either because competing underlying forms such as |im| do not exist in her lexicon, or because the semantic, pragmatic or syntactic components tell her top-down that a morpheme with an underlying form like |im| is ruled out by the context. Tableau (5), then, depicts a child that is given the underlying-phonetic pair |in+pa| [impa].

In tableau (5) we can see that the given pair |in+pa| [impa] does not match the child's winning triplet |in+pa| /.in.pa./ [inpa]: the phonetic part is different. From this discrepancy, the child will be able to learn.

In order to be able to learn, the child will have to choose a candidate in (5) as the one that she considers "correct". This "correct" candidate will have to be one that contains the given phonetic form [impa]. There are two candidates in (5) that meet this criterion: the second and the third candidate. The question now is: which of these two candidates will the child choose as the "correct" candidate? Following a general OT strategy, the child will choose the one that is the more harmonic of the two. This more harmonic candidate is the third: it violates only the middle ranked IDENTPLACE(|n|), whereas its sole competitor, the second candidate, violates the top-ranked */n/[m]. The "correct" candidate is denoted in (5) with a check mark.

This way of determining the "correct" candidate from among several candidates with different "hidden" phonological surface structures was originally proposed by Tesar (1997) and Tesar & Smolensky (2000) for a case in metrical phonology; their ROBUST INTERPRETIVE PARSING is identical in all respects to the procedure described here except for the way the mappings are computed: Tesar & Smolensky used only structural constraints, so that their mappings from Surface Form to Phonetic Form (their "Overt Form") and from Surface Form to Underlying Form were trivial (basically some erasing of structure). The present paper uses the complementary set of constraints (faithfulness, cue, articulatory), so that the relations between Surface Form and Phonetic Form and between Surface Form and Underlying Form are handled by the grammar.

When the child's own winner ("☞") and the candidate she considers correct ("√") are different, the child can take action. According to the gradual learning algorithm for Stochastic Optimality Theory (Boersma 1997, Boersma & Hayes 2001), she does this by raising the ranking of all constraints that prefer the "correct" candidate over the child's own winning candidate, and, conversely, by lowering the ranking of all constraints that prefer the child's own winner to the candidate she considers correct. In (5), therefore, we have to compare the constraint violations in the first candidate with those in the third. The only constraint that prefers the third ("correct") candidate to the first is *[TONGUETIP], so this is the only constraint that will rise; the only constraint that prefers the first (child-optimal) candidate to the third is IDENTPLACE(|n|), so this is the only constraint that will fall. In (5), the movements of the constraints are depicted by arrows. The constraint */n/[m] does not move, because it cannot choose between the first and the third candidate.

Given a language environment from which the child receives underlying-phonetic pairs, the learning algorithm in (5) will change the rankings of the constraints in the child's grammar. The following section investigates where the child will end up.

2 Validation of the learning procedure: mimicking your parents

Under some circumstances, children with the learning algorithm in (5) will mimic their parents' language. This section investigates a hypothetical language, and several types of children who try to learn it. The purpose of this section is to establish that the learning procedure in (5) is feasible; an application of the learning procedure to asymmetries in frequency and cue quality follows in section 3.

2.1 A language environment for coronal nasal place assimilation

For all the simulations in this section, I assume a language with eight morphemes, which can be connected to yield eight different sentences.

All sentences in this language consist of a sequence of two morphemes. The first morpheme is |in|, |et|, |um|, or |op|. The second morpheme is |pa|, |ma|, |ta|, or |na|. Because of some syntactic restrictions, only the following eight sentences can be made: |in+pa|, |in+ta|, |et+ma|, |et+na|, |um+ta|, |um+pa|, |op+na|, and |op+ma|.

I assume a Dutch or Catalan type of place assimilation: underlying |in+pa| is pronounced (by the parents) as [impa] in 80% of the cases, and as [inpa] in 20% of the cases, while underlying |et+ma| is pronounced as [etma] in 99% of the cases, and as [epma] in 1% of the cases. Underlying forms with labial codas are pronounced fully labially, and onsets do not assimilate to preceding codas. The full DISTRIBUTION of underlying-phonetic pairs is then given in (6).

(6) *Distribution of underlying-phonetic pairs in a Dutch-like assimilating language*

<table>
<tr><td>|in+pa| [inpa]</td><td>20</td><td>|um+ta| [umta]</td><td>100</td></tr>
<tr><td>|in+pa| [impa]</td><td>80</td><td>|um+ta| [unta]</td><td>0</td></tr>
<tr><td>|et+ma| [etma]</td><td>99</td><td>|op+na| [opna]</td><td>100</td></tr>
<tr><td>|et+ma| [epma]</td><td>1</td><td>|op+na| [otna]</td><td>0</td></tr>
<tr><td>|in+ta| [inta]</td><td>100</td><td>|um+pa| [umpa]</td><td>100</td></tr>
<tr><td>|in+ta| [imta]</td><td>0</td><td>|um+pa| [unpa]</td><td>0</td></tr>
<tr><td>|et+na| [etna]</td><td>100</td><td>|op+ma| [opma]</td><td>100</td></tr>
<tr><td>|et+na| [epma]</td><td>0</td><td>|op+ma| [otma]</td><td>0</td></tr>
</table>

In (6), the numbers are relative frequencies. Beside expressing the incidence of assimilation in coronal codas, the numbers also indicate that all eight underlying sentences are equally common, since for each underlying form the frequencies add up to a value of 100.

The following subsections report on computer simulations in which I feed this distribution to several types of learners, who differ in their constraint sets. What is the same across all these learners is their initial state: they all start out having all their constraints ranked at the same height, namely 100.0, and are then fed 400,000 pieces of language data (underlying-phonetic pairs). All learners have the grammar model of Stochastic OT (Boersma 1997, Boersma & Hayes 2001): at each evaluation of an incoming language datum, the learner chooses for each constraint a ranking randomly drawn from a Gaussian distribution around the ranking value of that constraint. The standard deviation of this EVALUATION NOISE is held constant at 2.0. In case an incoming datum leads to learning (i.e. if the learner's winner and the "correct" candidate are different), the change in the ranking of a moving constraint (the PLASTICITY) is a value that itself changes with the "age" of the learner: during the first 100,000 data, the plasticity is 1.0, during the next 100,000 data it is 0.1, during the next 100,000 data it is 0.01, and during the last 100,000 data it is 0.001. This plasticity scheme allows fast learning for young learners and accurate learning for older ones.

The first question to pose for every simulation is: will the learners mimic the distributions in (6)? That is, will they assimilate place in |in+pa| 80 percent of the time, and never in |um+ta|, just like their parents? The second question is: where will they localize this assimilation, in their phonology or in their phonetics? In other words, will they posit triplets like |in+pa| /.im.pa./ [impa] (phonological assimilation) or |in+pa| /.in.pa./ [impa] (phonetic assimilation), or both? This second question is an instance of a more general interesting issue, namely the construction of hidden representations; like other surface structures (e.g. foot structure), the surface forms in our example are inaudible and are not part of the primary language data; they have to be constructed by the learner. For the learning pair |in+pa| [impa], the child could probably construct either /.im.pa./ or /.in.pa./ as the "intermediate" form.

2.2 Simulation: a single constraint for place faithfulness

Our first learner has four cue constraints (*/p/[t], */m/[n], */t/[p], */n/[m]), one simple articulatory constraint Lazy (violated by the articulatory distance between the coda and the onset, in some arbitrary units; see §3.3), and a single faithfulness constraint IdentPlace (henceforth abbreviated in tableaus as IdPl).

After 400,000 learning pairs drawn from the distribution in §2.1, this learner ends up with the ranking values above tableau (7). The tableau shows what happens if the learner evaluates the input |in+pa| with zero evaluation noise: her 'disharmonies' (ranking values + noise) equal the numbers above the tableau, and the output will be /.in.pa./ [impa]. However, with full evaluation noise, this output will occur in only 80 percent of the evaluations. In the remaining 20 percent of the evaluations, namely when 91.5 + noise exceeds 93.9 + noise, the constraint */n/[m] will outrank Lazy, so that /.in.pa./ [inpa] wins. The numbers at the right of the tableau are the output probabilities given the underlying form |in+pa| and were computed by running this underlying form through the stochastic grammar 100,000 times.

(7) *The end state of learning an assimilating language*

	105.9	105.3	102.6	100.9	93.9	91.5	
|in+pa|	*/p/ [t]	*/m/ [n]	IdPl	*/t/ [p]	Lazy	*/n/ [m]	
☞ /.in.pa./ [impa]						*	80%
/.in.pa./ [inpa]					*!**		20%
/.im.pa./ [impa]			*!				0%
/.im.pa./ [inpa]		*!	*		***		0%

The answer to the first question (§2.1), then, is that this learner indeed replicates her parents' distribution for the underlying form |in+pa|. The answer to the second question is that her surface form is always /.in.pa./; in other words, any assimilation of |n| is entirely phonetic.

The learner handles the remaining underlying forms equally well. Tableau (8) shows how she replicates the adult 99~1 variation for the underlying form |et+ma|. The small difference between the learner's incidence of [epma] (0.7%) and that of the adults (1%) is typical of computer simulations with a finite number of data and a finitely small plasticity (Boersma & Hayes 2001). Again, this learner's assimilation is purely phonetic, except for the 0.1% probability that IdentPlace is ranked lower than */t/[p] and Lazy during an evaluation, in which case the assimilation becomes phonological.

(8) *The end state of learning an assimilating language*

	105.9	105.3	102.6	100.9	93.9	91.5	
|et+ma|	*/p/ [t]	*/m/ [n]	IdPl	*/t/ [p]	Lazy	*/n/ [m]	
/.et.ma./ [epma]				*!			0.7%
☞ /.et.ma./ [etma]					***		99.2%
/.ep.ma./ [epma]			*!				0.1%
/.ep.ma./ [etma]	*!		*		***		0%

The learner also replicates the adults' non-assimilation for labials, as (9) illustrates.

(9) *No assimilation for labial nasals and plosives*

	105.9	105.3	102.6	100.9	93.9	91.5	
|um+ta|	*/p/ [t]	*/m/ [n]	IdPl	*/t/ [p]	Lazy	*/n/ [m]	
/.un.ta./ [umta]			*!		***	*	0%
/.un.ta./ [unta]			*!				0.1%
☞ /.um.ta./ [umta]					***		99.9%
/.um.ta./ [unta]		*!					0%

	105.9	105.3	102.6	100.9	93.9	91.5	
|op+na|	*/p/ [t]	*/m/ [n]	IdPl	*/t/ [p]	Lazy	*/n/ [m]	
/.ot.na./ [opna]			*!	*	***		0%
/.ot.na./ [otna]			*!				0.1%
☞ /.op.na./ [opna]					***		99.9%
/.op.na./ [otna]	*!						0%

The remaining four underlying forms are produced entirely faithfully. As (10) illustrates for |in+ta|, all competing candidates are harmonically bounded.

(10) *No variation if underlying places are identical*

	105.9	105.3	102.6	100.9	93.9	91.5	
|in+ta|	*/p/ [t]	*/m/ [n]	IdPl	*/t/ [p]	Lazy	*/n/ [m]	
/.in.ta./ [imta]					*!**	*	0%
☞ /.in.ta./ [inta]							100%
/.im.ta./ [imta]			*!		***		0%
/.im.ta./ [inta]		*!	*				0%

The present section has shown that Stochastic OT learners can indeed infer intermediate hidden representations (in this case phonological surface structures) for this simple instance of place assimilation.[1]

The cause why all assimilation observed in this section is “phonetic”, i.e. due to the violation of cue constraints, is not a property of the learning procedure. Instead, it is a property of the constraint set: the factorial typology of the constraints does not allow phonological assimilation to be restricted to coronal nasals. This is because the only faithfulness constraint I considered is IdentPlace, a general constraint that cannot distinguish between nasals and plosives, nor between coronals and labials, and has therefore no hope of accounting for the Dutch type of place assimilation. The following section introduces a more granular account of faithfulness, so that phonological assimilation becomes possible.

2.3 Simulation: Jun’s (1995) granularity

To achieve more “phonological” solutions, this section implements a possibility inspired by Jun (1995), namely to have two separate faithfulness constraints for the two values of the place feature, namely IdentPlace(|labial|) and IdentPlace(|coronal|), and to have two separate faithfulness constraints for the two different manners, namely IdentPlace(nasal) and IdentPlace(plosive). With these constraints, the learner again replicates her parents’ distribution, as illustrated in (11).

(11) *Learning phonological assimilation*

	107.0	106.9	102.6	102.3	101.7	94.8	93.8	93.5	88.6	
|in+pa|	*/m/ [n]	*/p/ [t]	IdPl (|lab|)	*/t/ [p]	IdPl (plos)	Lazy	IdPl (nas)	*/n/ [m]	IdPl (|cor|)	
☞ /.in.pa./ [impa]								*		45%
/.in.pa./ [inpa]						*!**				19%
☞ /.im.pa./ [impa]							*		*	36%
/.im.pa./ [inpa]	*!					***	*		*	0%

---

[1] That hidden surface structure can be inferred in this way was shown for instances of metrical phonology by Tesar (1997) and Tesar & Smolensky (2000) for nonstochastic OT and by Boersma (2003) for stochastic OT. That hidden *underlying* structure can be inferred in a similar way was shown by Apoussidou (2007: ch.6). These works generally acknowledge that there are instances of languages that can be represented by the grammar model but cannot be learned by the learning procedure (see also §2.5).

In (11), 81 percent of the outputs have the phonetic form [impa], just as in the adult language. This time, however, the learner divides up her assimilation almost evenly between phonological (candidate 3: 36/81 = 44%) and phonetic (candidate 1: 45/81 = 56%).

2.4 Simulation: fully granular faithfulness

The next simulation follows Boersma (1998: 217) in having a separate IdentPlace constraint for each of the four underlying segments. In this way, the granularity of the faithfulness constraints corresponds exactly to that of the cue constraints, of which there is also one per (surface) segment. Tableau (12) shows the outcome.

(12) *A learner who varies between phonological and phonetic assimilation*

<table>
<tr><td></td><td>105.9</td><td>105.6</td><td>105.5</td><td>104.9</td><td>100.4</td><td>100.4</td><td>93.2</td><td>92.1</td><td>92.0</td><td></td></tr>
<tr><td>|in+pa|</td><td>IdPl (|m|)</td><td>*/p/ [t]</td><td>*/m/ [n]</td><td>IdPl (|p|)</td><td>IdPl (|t|)</td><td>*/t/ [p]</td><td>Lazy</td><td>*/n/ [m]</td><td>IdPl (|n|)</td><td></td></tr>
<tr><td>☞ /.in.pa./ [impa]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>*</td><td></td><td>40%</td></tr>
<tr><td>/.in.pa./ [inpa]</td><td></td><td></td><td></td><td></td><td></td><td></td><td>*!**</td><td></td><td></td><td>20%</td></tr>
<tr><td>☞ /.im.pa./ [impa]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>*</td><td>40%</td></tr>
<tr><td>/.im.pa./ [inpa]</td><td></td><td></td><td>*!</td><td></td><td></td><td></td><td>***</td><td></td><td>*</td><td>0%</td></tr>
</table>

The cue constraint */n/[m] and the faithfulness constraint IdentPlace(|n|) end up being ranked equally high. As a result, the surface form for the combination of underlying |in+pa| and phonetic [impa] will therefore be /.in.pa./ half of the time ("phonetic assimilation") and /.im.pa./ half of the time ("phonological assimilation"). The same 50-50 variation is seen for the 1 percent assimilation in underlying |t|, as (13) shows.

(13) *A learner who varies between phonological and phonetic assimilation*

<table>
<tr><td></td><td>105.9</td><td>105.6</td><td>105.5</td><td>104.9</td><td>100.4</td><td>100.4</td><td>93.2</td><td>92.1</td><td>92.0</td><td></td></tr>
<tr><td>|et+ma|</td><td>IdPl (|m|)</td><td>*/p/ [t]</td><td>*/m/ [n]</td><td>IdPl (|p|)</td><td>IdPl (|t|)</td><td>*/t/ [p]</td><td>Lazy</td><td>*/n/ [m]</td><td>IdPl (|n|)</td><td></td></tr>
<tr><td>/.et.ma./ [epma]</td><td></td><td></td><td></td><td></td><td></td><td>*!</td><td></td><td></td><td></td><td>0.5%</td></tr>
<tr><td>☞ /.et.ma./ [etma]</td><td></td><td></td><td></td><td></td><td></td><td></td><td>***</td><td></td><td></td><td>99%</td></tr>
<tr><td>/.ep.ma./ [epma]</td><td></td><td></td><td></td><td></td><td>*!</td><td></td><td></td><td></td><td></td><td>0.5%</td></tr>
<tr><td>/.ep.ma./ [etma]</td><td></td><td>*!</td><td></td><td></td><td>*</td><td></td><td>***</td><td></td><td></td><td>0%</td></tr>
</table>

This is a case, then, of a learner who varies between the two kinds of assimilation.

Other types of outcomes also exist. Tableau (14) shows a learner who ends up with a ranking that makes her a fully-phonetic assimilater, at least for the nasals.

(14) *Phonetic assimilation (for the nasals)*

	105.3	105.0	103.9	103.4	101.9	99.8	98.9	92.0	89.7	
|in+pa|	IDPL (|p|)	*/m/ [n]	*/p/ [t]	IDPL (|m|)	IDPL (|n|)	*/t/ [p]	IDPL (|t|)	LAZY	*/n/ [m]	
☞ /.in.pa./ [impa]									*	80%
/.in.pa./ [inpa]								*!**		20%
/.im.pa./ [impa]					*!					0%
/.im.pa./ [inpa]		*!			*			***		0%

What we see here is that there is variation not only *within* learners, but also *between* learners: different learners may end up in different grammars (constraint rankings), although their phonetic outputs are identical, so that there is no simple way to tell them apart just by listening to them. The message that linguists can take home is that they need not be too concerned if they find themselves disagreeing with other linguists on the exact analysis of a phonological and/or phonetic phenomenon: the disagreement may well reflect a divide that exists in real speech communities as well, or may indeed exist within single speakers.

2.5 Conclusion: the learning procedure is feasible

In sections 2.2 through 2.4 we have seen successful applications of the learning procedure in (5): the learners ultimately came to mimic the (conditional) probabilities of the phonetic forms that they encountered in their language environment, and they succeeded in positing reasonable phonological surface structures. As for the analysis of assimilation (such as its locus), they turned out to vary much in the same way as linguists tend to vary in their analyses of phonological phenomena.

It must be said here that the success of a few example learners does not guarantee that the learning procedure will always succeed if a constraint ranking exists. In fact, Tesar & Smolensky (2000) showed that their learning procedure with robust interpretive parsing and constraint demotion for nonstochastic OT failed to learn a metrical phonology for about 36 percent of constructed languages; likewise, Boersma (2003) showed that the present, quite similar, learning procedure for stochastic OT still failed to learn about 22 percent of Tesar & Smolensky's languages. On those same languages, even the same learning procedure for noisy Harmonic Grammar fails for about 12 percent (Boersma & Pater 2008). Future investigations will have to tell whether the learning procedure is incorrect or whether it fails in the same way as human learners do. For the moment, this learning procedure is the best we have, and I use it in the next section to establish two acquisitional biases in the ranking of faithfulness constraints, with the ultimate result of explaining markedness effects and licensing by cue.

3 Learning place assimilation from frequency and cue biases

Typologically, the ranking of faithfulness constraints has been observed to depend on two major circumstances:

(15) *Proposed universal rankings of faithfulness constraints*

a. **Ranking by environment**:
   1. *Positional faithfulness* (Beckman 1998):
      requires either an innately biased constraint set or an innate ranking.
   2. *Licensing by cue* (Steriade 1995):
      requires knowledge of auditory distances between phonological elements.

b. **Ranking by relative frequency**:
   Markedness by faithfulness (Boersma 1998: 180–184):
      requires knowledge of confusion probabilities between phonological elements.

The present section shows that these types of universal rankings emerge automatically as a side effect of learning, and that no innately biased constraint set, no innate ranking, and no extralinguistic knowledge of phonetic distances or confusion probabilities is required. Section 4 then relates these results to the earlier proposals in the literature, basically saying that Beckman and Steriade were correct in observing universal rankings of faithfulness, and that Boersma was correct in relating these to minimization of confusion, but that none of the three provided the correct underlying mechanism.

3.1 Generation 1: perfectly faithful, and with a frequency bias

In order to prove that the universal rankings in (15) emerge automatically as a side effect of acquisition in a noisy world, I have to start by positing a language in which none of these universal rankings hold, and to proceed by feeding this language to a fresh group of virtual learners.

The Generation 1 that I propose, then, is a group of speakers who have none of the above-mentioned universal rankings. The easiest way to accomplish this is to assume that they have no faithfulness violations and also implement the surface forms with perfect articulations. In order for the second generation to have some asymmetries to work with, this first generation does have nasals, plosives, labials, and coronals. In fact, they have the exact same eight morphemes and eight sentences as the speakers discussed in §2.1. The speakers also have a CORONAL FREQUENCY BIAS of 2.0, i.e. underlying coronal codas occur twice as often as underlying labial codas. This corresponds to a language like English, in which coronal codas are about three times more frequent than labial codas and about four times more frequent than velar codas, as a quick count of a written text shows. The speakers' production distribution can therefore be described with the triplets in (16).

(16) *Distribution of underlying-surface-phonetic triplets in a perfectly faithful language*

|et+ma| /.et.ma./ [etma]	200	|op+ma| /.op.ma./ [opma]	100
|op+na| /.op.na./ [opna]	100	|et+na| /.et.na./ [etna]	200
|in+pa| /.in.pa./ [inpa]	200	|um+pa| /.um.pa./ [umpa]	100
|um+ta| /.um.ta./ [umta]	100	|in+ta| /.in.ta./ [inta]	200

3.2 Input for Generation 2: transmission noise

The perfect phonetic output of Generation 1 in (16) is changed by the imperfections of the transmission channel (background noises produced by wind, other speakers, and blood flow). For the present simulation, I simplifyingly assume that the transmission noise does not affect the auditory cues of the onset consonants of the second syllable, which are therefore always the "perfect" [m], [n], [p], and [t]. I also assume that it does not affect the codas of the first syllable that are underlyingly homorganic to the following onsets (the cases in the second column of (16)).

The transmission noise affects the heterorganic codas (those in the first column of (16)) by changing their auditory properties in a gradual way. Specifically, the transmission noise does not necessarily change a perfect [n] into a perfect [m] or the reverse. Real possibilities for auditory place are not just [n] and [m], but a large number of values along an auditory place continuum, most likely of a spectral nature. In this paper I simply assume a discretization of the place continuum into four values: the continuum for nasals from labial to coronal is [m] – [M] – [N] – [n], where [M] and [N] are a "reduced" [m] and [n], respectively, and the continuum for plosives is [p] – [P] – [T] – [t]. The transmission noise, then, may change the adult [n] into something that the learner will hear as [n], [N], [M], or [m].

Importantly, the relative influence of "noise" is larger if the "signal" (the auditory cues for place) is weaker. Nasals such as [m] and [n] share a large part of their spectral properties, namely the resonance of the nasal tract and the fact that the oral cavity acts as a sidebranch that causes a spectral "zero" (valley). The distinction between [m] and [n] therefore lies mainly in the formant movements around the oral closure (Malécot 1956). Plosives, by contrast, have not only these formant movements, but also contain place information during the oral release burst, so that they are likely to have more and/or better place cues than nasals. This has been confirmed in perceptual discrimination experiments (Mohr & Wang 1968, Pols 1983). For the present paper, then, I assume that the auditory place values in the learner's input data are more variable for nasals than for plosives.

Taking into account both the coronal frequency bias and the greater influence of transmission noise for nasals, I can describe the primary language data (pairs of underlying and phonetic form) for a learner of generation 2 as the distribution in (17), which I based on two Gaussian distributions.

(17) *Distribution of language data for a learner of a perfectly faithful language*

|et+ma|	[epma]	2	|op+ma|	[opma]	100
|et+ma|	[ePma]	16	|op+ma|	[oPma]	0
|et+ma|	[eTma]	68	|op+ma|	[oTma]	0
|et+ma|	[etma]	114	|op+ma|	[otma]	0
|op+na|	[opna]	57	|et+na|	[epna]	0
|op+na|	[oPna]	34	|et+na|	[ePna]	0
|op+na|	[oTna]	8	|et+na|	[eTna]	0
|op+na|	[otna]	1	|et+na|	[etna]	200
|in+pa|	[impa]	22	|um+pa|	[umpa]	100
|in+pa|	[iMpa]	44	|um+pa|	[uMpa]	0
|in+pa|	[iNpa]	62	|um+pa|	[uNpa]	0
|in+pa|	[inpa]	72	|um+pa|	[unpa]	0
|um+ta|	[umta]	36	|in+ta|	[imta]	0
|um+ta|	[uMta]	31	|in+ta|	[iMta]	0
|um+ta|	[uNta]	22	|in+ta|	[iNta]	0
|um+ta|	[unta]	11	|in+ta|	[inta]	200

The numbers in (17) are again relative token frequencies. The coronal frequency bias is evident from the fact that underlying forms with a coronal coda in their first syllable have a total frequency of 200, whereas underlying forms with a labial coda have a total frequency of only 100. In the first column in (17) we can see that plosives have relatively good cues: in the presence of a heterorganic coda, plosives have perfect auditory cues in 57 percent of the cases; in (17) this is true of both |op+na| and |et+ma| (57% of 200 is 114); we can also see that the place value of an underlying plosive turns into its perfect opposite only 1 percent of the time. By contrast, nasals have relatively poor cues: the cues are perfect in only 36 percent of the cases, for both |um+ta| and |in+pa| (36% of 200 is 72), and the place value of an underlying nasal turns into its perfect opposite no less than 11 percent of the time. The second column of (17) expresses the somewhat artificial assumption that for homorganic coda-onset sequences the intended auditory place of the coda is perfectly audible.

3.3 The learner: arbitrary cue constraints

With the four possible values of auditory place needed to describe the distribution in (17), the learner needs more cue constraints than just the four that the learners of §2 could do with. For the surface phonological element /n/, for instance, the learner needs the following four cue constraints: */n/[n], */n/[N], */n/[M], and */n/[m]; for /p/, she needs */p/[p], */p/[P], */p/[T], and */p/[t]. In total, the learner needs 16 cue constraints.

The character of the set of cue constraints has changed between §2 and the present section. In §2, each cue constraint militated against a phonetic implementation of the surface segment as the "opposite" phonetic segment; in other words, the set of cue constraints contained a bias towards an "identity" between surface form and phonetic form (I write "identity" here with scare quotes, because one cannot really speak of "identity" in comparing a discrete phonological structure, e.g. binary feature values, with a continuous auditory representation, e.g. a basilar excitation pattern). In the present section, the set of cue constraints have no such "identity" bias: the presence of both */n/[m] and */n/[n], for instance, means that as far as the total set of cue constraints is concerned, the auditory cues are arbitrarily related to the phonological segments, and that it is the learner's ultimate *ranking* of the cue constraints that will have to establish that /n/ relates better to [n] than to [m].

This arbitrariness of the cue constraints is a common starting point in work on perception in OT (Escudero & Boersma 2003, 2004; Boersma & Hamann 2007a). For establishing a correct typology it relies heavily on the learning procedure and on the availability of a sufficiently rich set of language data. In this respect it is crucial for the present learner that the language data contain underlyingly homorganic clusters, i.e. the second column in (17).

Apart from the 16 cue constraints, the learner has the same four faithfulness constraints as in §2.4 and the same articulatory constraint as in §§2.2–4. For simplification, I assume that the degree of violation of LAZY corresponds to the auditory distance between the two consonants in the cluster. That is, I assume that [ampa] causes zero violations of LAZY, [aMpa] causes 1 violation, [aNpa] causes 2 violations, and [anpa] causes 3 violations (in the child's idea of how to pronounce this). A more intelligent approximation of the relation between articulatory effort and constraint violation is probably possible, but this would not really influence the result.

3.4 The simulation procedure

This time the simulation does not involve just a couple of virtual learners, as in the simulations of §2, but one thousand virtual learners. Each of these learners has the same four faithfulness constraints, 16 cue constraints, and one articulatory constraint, and each learner starts with all 21 constraints ranked at 100.0. Each learner is subsequently fed with 400,000 underlying–phonetic pairs randomly selected from the distribution in (17), with the same decreasing plasticity as discussed in §2.1 (1.0~0.1~0.01~0.001), and the same evaluation noise (standard deviation 2.0). The reason for simulating so many learners is that learners are expected to vary in their learning paths, as a result of the different random orders of data that they are fed, and the different stochastic rankings that they use during evaluations.

3.5 Simulation result: the grammar of the median learner

The ultimate ranking values of the MEDIAN LEARNER are given in (18). The median learner is a hypothetical learner, not one of the thousand simulated learners. For each constraint, this learner has a ranking that is the median of the rankings for the 1000 learners.

(18) *Ultimate rankings of the median learner*

*/t/[p]	108.2	*/m/[N]	99.6
*/p/[t]	107.5	*/n/[N]	99.6
*/n/[m]	104.5	*/t/[T]	99.2
IdPl(|p|)	103.7	Lazy	98.6
*/m/[n]	103.2	*/p/[P]	98.5
*/t/[P]	103.2	*/m/[M]	98.0
IdPl(|t|)	102.9	*/m/[m]	97.3
*/p/[T]	102.3	*/n/[n]	96.4
IdPl(|m|)	101.9	*/p/[p]	91.0
*/n/[M]	101.4	*/t/[t]	90.4
IdPl(|n|)	100.0		

3.6 The production behaviour of the median learner

The median learner does not replicate the relative frequencies of the adults in (16), nor her incoming distribution in (17). The tableaus in (19) through (23) show what she does do when asked to produce the eight underlying sentences. For reasons of space, each tableau includes only a subset of the 21 constraints, namely those that are violated in one or more candidates in that tableau; the reader can verify that the ranking values above each tableau are identical to the corresponding ones in (18). The percentages after each tableau give the probabilities of each of the eight possible outputs, measured by running the underlying form through her grammar 10,000,000 times (if we instead run the underlying form through every "real" virtual learner's grammar 10,000 times, we get very similar results).

(19) *The production behaviour of the median learner*

	104.5	103.2	101.4	100.0	99.6	99.6	98.6	98.0	97.3	96.4	
|in+pa|	*/n/ [m]	*/m/ [n]	*/n/ [M]	IDPL (|n|)	*/m/ [N]	*/n/ [N]	LAZY	*/m/ [M]	*/m/ [m]	*/n/ [n]	
i. /.in.pa./ [impa]	*!										0.7%
ii. /.in.pa./ [iMpa]			*!				*				7.0%
iii. /.in.pa./ [iNpa]						*!	**				17.8%
☞ iv. /.in.pa./ [inpa]							***			*	43.2%
v. /.im.pa./ [impa]				*!					*		28.0%
vi. /.im.pa./ [iMpa]				*!			*	*			2.8%
vii. /.im.pa./ [iNpa]				*!	*		**				0.4%
viii. /.im.pa./ [inpa]		*!		*			***				0.0%

For underlying |in+pa|, we can see that the learners have a special bias towards the least-assimilating form [inpa] (43.2%, whereas only 36% in the input) and towards the most-assimilating form [impa] (28.7% instead of 11%). If full assimilation ([impa]) occurs, it is usually phonological (candidate v: 28%) rather than phonetic (candidate i: 0.7%).

(20) *The production behaviour of the median learner*

	108.2	107.5	103.2	102.9	102.3	99.2	98.6	98.5	91.0	90.4	
|et+ma|	*/t/ [p]	*/p/ [t]	*/t/ [P]	IDPL (|t|)	*/p/ [T]	*/t/ [T]	LAZY	*/p/ [P]	*/p/ [p]	*/t/ [t]	
i. /.et.ma./ [epma]	*!										0.02%
ii. /.et.ma./ [ePma]			*!				*				3.9%
iii. /.et.ma./ [eTma]						*!	**				33.7%
☞ iv. /.et.ma./ [etma]							***			*	55.7%
v. /.ep.ma./ [epma]				*!					*		6.6%
vi. /.ep.ma./ [ePma]				*!			*	*			0.0%
vii. /.ep.ma./ [eTma]				*!	*		**				0%
viii. /.ep.ma./ [etma]		*!		*			***				0%

For |et+ma| in (20), we see that non-assimilation ([etma]) and minimal assimilation ([eTma]) occur with frequencies comparable to those in (17); [eTma] occurs exclusively as phonetic assimilation. Full assimilation ([epma]) again occurs more often in (20) than in (17), and is nearly exclusively phonological (candidate v): full phonetic assimilation (candidate i) occurs in only 0.02% of the cases. Conversely, partial assimilations ([eTma], [ePma]) are virtually always phonetic, i.e. have a phonological surface form /.et.ma./: the candidates where a phonological assimilation is partially or wholly "undone" in the phonetics (vi, vii, viii) are very rare: only candidate vi occurs 4 times (out of 10,000,000). When comparing (20) with (19), we see that the plosive tends to be produced much less variably than the nasal, as expected.

(21) *The production behaviour of the median learner*

	108.2	107.5	103.7	103.2	102.3	99.2	98.6	98.5	91.0	90.4	
|op+na|	*/t/ [p]	*/p/ [t]	IDPL (|p|)	*/t/ [P]	*/p/ [T]	*/t/ [T]	LAZY	*/p/ [P]	*/p/ [p]	*/t/ [t]	
i. /.ot.na./ [opna]	*!		*				***				0%
ii. /.ot.na./ [oPna]			*!	*			**				0%
iii. /.ot.na./ [oTna]			*!			*	*				0%
iv. /.ot.na./ [otna]			*!							*	3.7%
v. /.op.na./ [opna]							***!		*		45.8%
☞ vi. /.op.na./ [oPna]							**	*			42.2%
vii. /.op.na./ [oTna]					*!		*				8.3%
viii. /.op.na./ [otna]		*!									0.06%

For the labial plosive in |op+na|, the learners fully assimilate in only 3.8% of the cases, which is less than for the coronal plosive in (20). If they do assimilate fully, they again strongly prefer to do so in the phonology (candidate iv) rather than in the phonetics (candidate viii); if they assimilate partially, they again virtually always do that in the phonetics (candidates vi and vii), not in the phonology (candidates ii and iii: zero attestations in 10 million tokens, though neither is harmonically bounded). By the way, the fact that the zero-noise winner ("☞" in the tableau) is not the most frequent winner when the noise is 2.0 (which is candidate v: 45.8%) is a normal situation in stochastic OT: candidates vii and iv draw more wins away from candidate vi than from candidate v.

(22) *The production behaviour of the median learner*

	104.5	103.2	101.9	101.4	99.6	99.6	98.6	98.0	97.3	96.4	
|um+ta|	*/n/ [m]	*/m/ [n]	IDPL (|m|)	*/n/ [M]	*/m/ [N]	*/n/ [N]	LAZY	*/m/ [M]	*/m/ [m]	*/n/ [n]	
i. /.un.ta./ [umta]	*!		*				***				0.0%
ii. /.un.ta./ [uMta]			*!	*			**				0.01%
iii. /.un.ta./ [uNta]			*!			*	*				0.14%
iv. /.un.ta./ [unta]			*!							*	11.4%
v. /.um.ta./ [umta]							***!		*		22.7%
☞ vi. /.um.ta./ [uMta]							**	*			32.8%
vii. /.um.ta./ [uNta]					*!		*				28.9%
viii. /.um.ta./ [unta]		*!									4.1%

For |um+ta|, we find the usual features: if assimilation is full, the learners have a preference for performing the assimilation in the phonology (11.4%) over doing it in the phonetics (4.1%), but if the assimilation is partial, they strongly prefer to do it in the phonetics (candidates vi and vii: 61.7%) rather than partially "reversing" the phonology (candidates ii and iii: 0.15%). The total incidence of full assimilation for these labial nasals (15.5%) is more than for the labial plosives of (21) (3.8%) but less than for the coronal nasals in (19) (28.7%). Together with the incidence of full assimilation for the coronal plosives of (20) (6.6%), these full assimilation counts probabilistically reflect the typology in (1). This suggests that (1a) emerges from a frequency difference between coronals and other articulators, and that (1b) emerges from cue reliability differences between nasals and plosives.

To be complete, I discuss briefly how the learners handle the underlyingly homorganic sequences. Tableau (23) shows the production of the underlying form |in+ta|.

(23) *The production behaviour of the median learner*

	104.5	103.2	101.4	100.0	99.6	99.6	98.6	98.0	97.3	96.4	
|in+ta|	*/n/ [m]	*/m/ [n]	*/n/ [M]	IDPL (|n|)	*/m/ [N]	*/n/ [N]	LAZY	*/m/ [M]	*/m/ [m]	*/n/ [n]	
i. /.in.ta./ [imta]	*!						***				0.03%
ii. /.in.ta./ [iMta]			*!				**				1.0%
iii. /.in.ta./ [iNta]						*!	*				5.3%
☞ iv. /.in.ta./ [inta]										*	90.4%
v. /.im.ta./ [imta]				*!			***!		*		1.1%
vi. /.im.ta./ [iMta]				*!			**	*			1.2%
vii. /.im.ta./ [iNta]				*!	*		*				0.7%
viii. /.im.ta./ [inta]		*!		*							0.12%

For |in+ta| we see a large preference for the fully faithful surface-phonetic pair /.in.ta./ [inta]. The only substantial assimilation is phonetic and minimal (candidate iii). Quite similar results are obtained for the underlying sentences |et+na|, |op+ma|, and |um+pa|.

3.7 The effects of frequency and cue reliability on the ranking of faithfulness

If we look at just the phonological assimilations (the last four candidates in each of (19) and (20), and the first four candidates in each of (21) and (22)), we see that they occur most for |n| (31.2%), somewhat less for |m| (11.6%) and |t| (6.6%), and least for |p| (3.7%). Just as the full assimilation counts in §3.6, these phonological assimilation counts correspond to the implicational universals in (1). It is important to note that these universals have emerged automatically within one generation of learners: the parents had no assimilation, but their children introduced it, and skewed it on the basis of frequency differences and of differences in cue quality.

When we study tableaus (19) through (22), we can easily find what causes the segment-dependency of the likelihood of phonological assimilation. The cause is the differential ranking of faithfulness constraints. The median learner, and also the great majority of the separate learners, have two kinds of faithfulness rankings that explain the universals.

One of the faithfulness rankings that we see in almost all learners is based on the frequency difference between labials and coronals: we see that faithfulness constraints tend to end up being ranked higher for the less common place value (labial) than for the more common place value (coronal). The numbers are summarized in (24).

(24) *Faithfulness ranked by frequency*

IdentPlace (|p|) >> IdentPlace (|t|)	(982 learners)
IdentPlace (|m|) >> IdentPlace (|n|)	(968 learners)

The other faithfulness rankings that we see in almost all learners is based on differences in auditory cue quality: we see that faithfulness constraints tend to end up being ranked higher in an environment with more reliable place cues (plosives) than in an environment with less reliable place cues (nasals). The numbers are in (25).

(25) *Faithfulness ranked by cue quality*

IdentPlace (|p|) >> IdentPlace (|m|)	(953 learners)
IdentPlace (|t|) >> IdentPlace (|n|)	(988 learners)

Not only the median learner, but 910 out of 1000 learners end up with the complete ranking that has been proposed by Boersma (1998: 217) on the basis of frequency and confusability (but without simulations), which is summarized in Fig. 3.

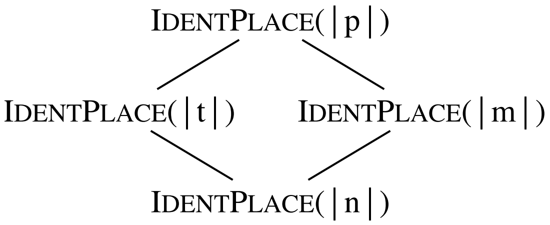

Figure 3

Faithfulness ranked by frequency and cue quality.

Figure 3 suggests that the authors mentioned at the beginning of §3 are correct in attributing the observed phenomena to faithfulness rankings. That is, Steriade (1995) is correct in expressing licensing-by-cue effects with faithfulness rankings, Beckman (1998) is correct in expressing positional faithfulness as faithfulness (rather than as markedness), and Boersma (1998) is correct in expressing (frequency-dependent) markedness in terms of faithfulness rankings. What all these authors miss, however, is the underlying mechanism that brought about these rankings. The present paper fills that gap.

3.8 The effects of frequency and cue reliability on the ranking of cue constraints

Not only the faithfulness constraints, but also the cue constraints tend to end up with useful rankings, as (18) shows and (26) summarizes.

(26) *Cue constraint rankings*

a. The farther your cue is removed from the “best” cue, the worse it is:
   */t/[p] >> */t/[P] >> */t/[T] >> */t/[t]
   */p/[t] >> */p/[T] >> */p/[P] >> */p/[p]
   */n/[m] >> */n/[M] >> */n/[N] >> */n/[n]
   */m/[n] >> */m/[N] >> */m/[M] >> */m/[m]

b. Adverse place cues are taken more seriously in plosives than in nasals:
   */t/[p] >> */n/[m]
   */p/[t] >> */m/[n]

c. Labial place cues are taken more seriously than coronal place cues:
*/t/[p] >> */p/[t]
*/n/[m] >> */m/[n]

However, since the present paper focuses on explaining the mechanism behind markedness and licensing by cue, the remaining sections discuss the rankings of faithfulness constraints alone.

4 Attested (near-)universal rankings of faithfulness

Many different types of (near-)universal rankings of faithfulness constraints have been proposed in the literature. What do these rankings reflect, and to what causes did the authors ascribe their universality?

4.1 Universal faithfulness rankings reflect phonological context

Beckman (1998) observes that certain phonological features are less likely to neutralize in some positions than in others. For instance, voicing neutralization is more likely in codas than in onsets. To account for this asymmetry, Beckman proposes that Universal Grammar contains a general constraint IDENTVOICE "an underlying voice specification should appear unchanged in the surface form" and a more specific constraint IDENTVOICE (onset) "an underlying voice specification should appear unchanged in the surface form, if the relevant segment is in onset".[2] The observed asymmetry is then predicted, because the constraint set itself cannot represent a language in which onsets neutralize for voice but codas do not.

Beckman (1998: 36) goes on to remark that without loss of generality, the two constraints can always be ranked in the order "specific >> general", as in (27).

(27) *Fixed ranking by specificity (Beckman 1998: 36)*
IDENTVOICE (onset) >> IDENTVOICE

Beckman subsequently proposes that the specific-over-general ranking in (27) is given as fixed by UG. If this is so, this would be empirically equivalent to proposing the fixed ranking in (28) (under the assumption that "onset" and 'coda" are the only possible positions).

(28) *Equivalent fixed ranking by position*
IDENTVOICE (onset) >> IDENTVOICE (coda)

While the reformulation in (28) loses the elegant automatic specific-over-general idea, it is in all empirical respects equivalent to the ranking in (27): it produces the same typology of neutralization.

In the context of the present paper, the ranking in (28) comes with a large advantage over (27): the ranking in (28) can be predicted to emerge automatically during acquisition. That is, *if* faithfulness constraints are positional (i.e. conditioned by syllable position, as in (28)), *and* auditory voicing cues are *on average* more reliable in onset than in coda, *then* the learner will come to rank IDENTVOICE (onset) above IDENTVOICE (coda), *even if* no such ranking is evident in the parents' productions.

---

[2] I adapted Beckman's notation to the one used in the present paper.

The constraint set in (28) places less of a burden on UG than the set in (27) does: the set in (27) assumes an innate bias disfavouring neutralization in onsets, whereas the set in (28) only assumes that constraints can be ranked by syllable position; the learning procedure will take care of the ranking. It is also imaginable that the constraints in (28) themselves could more easily emerge from language data than the constraint set in (27).

I conclude that the predictions of the present model are compatible with Beckman's account of positional faithfulness. When Beckman's proposals are trivially reformulated in the manner of (28), they are predicted by the learning procedure in (5).

4.2 Universal faithfulness rankings reflect cue audibility

Steriade (1995, 2001), basing herself on evidence about directional asymmetries in place assimilation of apicals, argues that positional faithfulness (or PROSODIC LICENSING) accounts such as that in §4.1 rely too much on prosodic positions such as onset and coda. According to Steriade, what is really going on in the case of §4.1 is that voicing neutralization is more likely in contexts where obstruent voicing cues are hard to hear, such as before consonants (especially obstruents) and word-finally, than in contexts where obstruent voicing cues are easy to hear, such as before vowels and other sonorants. Steriade translates this idea into a universal ranking of fatithfulness constraints, as in (29).

(29) *Ranking by cue (Steriade 1995)*

IMPLEMENT[shortVOT] / _[+son] >> IMPLEMENT[shortVOT] / _#

The conditioning in (29) is meant to be phonetic: the contexts [+son] and # (word boundary) are shorthand for "environments with many well-audible cues for [shortVOT]" and "environments with few and poor cues for [shortVOT]", respectively. In spirit, Steriade's model is more closely related to my model than Beckman's model is, but Steriade's model is more difficult to translate into my model, because it includes both phonological and phonetic representations.

The translation involves getting clear what the representations are. The arguments to the IMPLEMENT constraints are auditory features such as [NegativeVOT]. Steriade makes it explicit that the phonological surface form contains such auditory features; presumably (although Steriade does not make this explicit), the underlying form contains auditory features as well, because otherwise the idea of faithfulness would make no sense. According to Steriade, the surface form contains articulatory features as well, which are connected to the auditory features by "immutable physical facts" (p.1:18); for the auditory feature [shortVOT], for instance, they are the required articulatory feature [glottal adduction] (we could add [lung contraction]) and one or more of the articulatory features [larynx lowering], [short closure], [vocal cord slackening], and [oral cavity expansion]. Steriade further argues that auditory features must be distinguished from cues; for the auditory feature [shortVOT], she mentions the possible cue of closure voicing, and we could perhaps add burst amplitude and the actual VOT value. Steriade states that auditory features and cues "are correlated in some way" (p.1:26). From these considerations, we can establish Steriade's view of representations and their relations, as in (30).

(30) *Representations and their relations proposed by Steriade (1995)*

a. underlying-auditory ~ surface-auditory: related by faithfulness constraints;
b. surface-auditory ~ auditory cues: related "in some way";
c. surface-auditory ~ surface-articulatory: related by "immutable physical facts".

Although (30) contains four distinct entities, Steriade acknowledges only two representations: a surface form, with auditory and articulatory parts, and an underlying form. To translate (30) to the model used in the present paper, we have to add only one hypothesis, namely that *the cues have to be taken seriously as a representation*. From this one hypothesis, all else follows.

The first thing that follows from regarding cues as a representation, is that they become a serious competing candidate for the relation in (30c). After all, to which kind of representation are articulatory gestures such as [larynx lowering], [short closure], [vocal cord slackening], and [oral cavity expansion] most "immutably" related: to the auditory feature [shortVOT], which is "relevantly invariant" (p.1:25), or to the auditory cues [closure voicing], [burst amplitude], and [actual VOT]? I believe the relation is to the cues, especially if the context is given. That is, articulatory gestures are more immutably related (namely, by the speaker's acquired sensorimotor knowledge) to the actual sounds than to some relatively abstract invariant auditory property of it.

A thing that follows from replacing the "surface-auditory" representation in (30c) by auditory cues is that the "surface-auditory" representation that remains in (30a) and (30b) no longer has to be auditory. Instead, it can be replaced by a traditional abstract phonological representation such as [+voi]. The advantage of this is that one no longer has to look for an invariant auditory property of the feature, a quest that may often turn out to be elusive.

The second thing that follows directly from regarding cues as a representation, is that their relation to the surface-auditory features in (30b) has to be handled by the grammar. In fact, Steriade (p.1:26) explicitly states that this relation is specific to the language at hand: in English, closure voicing acts as a cue to [shortVOT], in French it does not (according to Steriade, it is a feature of its own).

With these three amendments, all derived from taking cues seriously as a representation, we can translate (30) into (31).

(31) *Representations and their relations*

a. underlying ~ surface: related by faithfulness constraints;
b. surface ~ auditory cues: related by cue constraints;
c. auditory cues ~ articulations: related by sensorimotor knowledge.

With (31) we have arrived at the model by Boersma (1998, 2007ab). For the present paper, the two representations in (31c) have been combined into one Phonetic Form, as in Fig. 1. In all, (31) contains one representation more than (30). In terms of (31), the faithfulness ranking for obstruent voicing would look as (32).

(32) *Indirect ranking by cue*

IDENT[+voi] / _[+son] >> IDENT[+voi] / _#

In going from (29) to (32), faithfulness constraints have remained, but their arguments are now the more traditional abstract phonological features, and their phonetic conditioning has turned into a more traditional phonological conditioning (by [+son] and #). In later work, Steriade (2001) works with constraints similar to those in (32);

she no longer considers auditory and articulatory representations, but instead ascribes the universality of rankings such as (32) to the speaker's knowledge of universal auditory distances between phonological entities (such as between [+voi] and [–voi] segments) in various contexts (the "P-map").

The present model predicts that rankings such as (32) automatically emerge during acquisition: *if* faithfulness constraints are as granularly context-dependent as Steriade (2001) proposes, *and* voicing cues are better before sonorants than word-finally, *then* the learner will come to rank IDENT[+voi]/_[+son] above IDENT[+voi]/_#, even if her parents do not.

I conclude that the predictions of the present model are fully compatible with Steriade's account of licensing by cue. However, the separate device of "P-map" is not needed to account for the rankings: they are predicted by the explicit learning procedure in (5). The present account can thus be said to corroborate Steriade's analyses by providing the mechanism underlying the cue-based rankings that she proposes.

That the P-map has become unnecessary for explaining the ranking of faithfulness, does not directly rule out the explanatory force of the P-map proposal in general. There is one independent linguistic process in which Steriade (2001: 238) sees evidence for the P-map: loanword adaptation. However, Boersma (2007b) and Boersma & Hamann (2007b) have shown that the loanword phenomena that Steriade describes can be straightforwardly accounted for by modelling perceptual loanword adaptation as a mapping from the phonetic to the surface form in Fig. 1, evaluated by structural and cue constraints. Thereby, no linguistic processes that require the P-map seem to remain.[3]

4.3 Universal faithfulness rankings reflect frequency of occurrence

The proposals by Beckman and Steriade concerned the dependence of faithfulness for the *same* feature value in *different* contexts, for example for the dependence of place faithfulness on position (onset, coda, prevocalic) or on manner (plosive, nasal). The ranking of faithfulness can also be different for *different* feature values in the *same* context; for example, place faithfulness can be ranked higher for labial place than for coronal place, in the context of nasality. Here we enter the area of MARKEDNESS.

Differences in the propensity of different feature values to undergo phonological changes has often been ascribed to markedness relations. Thus, the greater changeability of coronals as compared to labials has been ascribed to coronal unmarkedness (for an overview, see Paradis & Prunet 1991). Instead, Boersma (1998: 180–184) proposes a universal ranking of faithfulness as a function of frequency of occurrence, as in (30).

(30) *Ranking by frequency (Boersma 1998)*
IDENTPLACE (labial) >> IDENTPLACE (coronal)
(if [labial] is less frequent than [coronal])

Boersma ascribes this ranking to the idea that speakers are listener-oriented: if coronal codas are more frequent underlyingly than labial codas, listeners will expect more coronals than labials and optimize their perceptual boundaries in such a way that they obtain a bias towards recognizing underlying coronals (attested by Hura, Lindblom &

---

[3] Steriade (2001: 237–238) also mentions the paralinguistic process of the selection of rhymes in poetry. It is not clear what path through the representations such a process should take (from surface to auditory and back?), so I will refrain here from advancing a detailed alternative proposal in terms of Fig. 1.

Diehl 1992: 66), so that speakers can get away with changing underlying coronals more easily than with changing underlying labials (also Steriade 2001: 236).

The present model predicts that rankings such as (30) automatically emerge during acquisition: *if* coronals are more frequent in codas than labials are, *then* the learner will come to rank IDENTPLACE (labial) over IDENTPLACE (coronal) (for codas), even if her parents do not.

I conclude that the predictions of the present model are fully compatible with Boersma's account of markedness-by-faithfulness. However, the speaker does not have to explicitly take the listener into account: the learning procedure in (5) will take care of ranking the constraints appropriately.

A question that remains is what causes the frequency differences in the first place. These causes may be in the phonetics: perhaps coronals are easier to produce in codas than labials, or they are faster to produce so that they obscure the preceding vowels to a lesser extent; likewise, the reason why [−round] nonlow vowels are more frequent than [+round] nonlow vowels, both tokenwise (as any quick count of a Spanish text indicates) and in inventories (Maddieson 1984), may be because rounding tends to obscure auditory contrasts in height and place features. In general, any number of causes of a non-structural nature may influence the frequency of specific feature values (perhaps helped by the lexical selection mechanism discussed in §5.5). What is important is that once a frequency difference exists, the present model predicts a difference in the propensity to undergo change in the phonology.

The hypothesis that markedness-as-phonological-strength is indeed caused by markedness-as-frequency, makes empirical predictions. It predicts that in languages where the frequency relations happen to be reversed, the phonological strengths are also reversed. Boersma (1998: 224) mentions the case of Tagalog, where dorsal codas are as common as coronal codas (and much more common than labial codas), and indeed dorsal codas can undergo assimilation (though not as often as coronals), whereas labial codas cannot. A more convincing example is given by Hume & Tserdanelis (2002), who show that in Sri Lankan Portuguese Creole the labial |m| is nearly twice as frequent as the coronal |n| (in codas) and indeed is much more likely than |n| to undergo deletion and assimilation to the following consonant; while Hume & Tserdanelis do not propose a causal relationship between frequency and phonological strength (they still speak in terms of "markedness"), the present paper has provided such a relationship: the learning procedure in (5) predicts that frequency differences cause strength differences. Nevertheless, an apparent counterexample also seems to exist (Adam Albright, p.c.): in Korean, final coronals are rare yet undergo assimilation more easily than other places of articulation; a detailed analysis of this language would be extremely interesting.

The present paper cannot begin to address all the phenomena that have been ascribed to the concept of markedness since phonology emerged as a science. However, two major aspects of it, namely frequency and phonological strength, have now been related to one another by the learning algorithm in (5). Perhaps the remaining aspects of markedness can undergo similar fates; if so, Menn's (1983) hope that markedness can be dispensed with as an explanatory concept, will have come true.

4.4 Conclusion

Although all previously proposed universal faithfulness rankings can be said to have their distal and/or proximal causes in the phonetics, all the proposed explanations can be dispensed with. Thus, we can dispense with innate ranking (e.g. by position), extralinguistic knowledge (e.g. the P-map), and listener-oriented Bayesian computation (of expectation biases). The learning procedure in (5) predicts the proposed rankings.

The account presented here extends to all areas of phonology that occupy themselves with substantive features (rather than, say, metrical structure). As for ranking by cue, the account applies not only to the cases of context-based place assimilation and voice neutralization discussed here, but potentially to all cases of assimilation and neutralization. As for markedness-by-faithfulness, the account applies not only to the frequency biases for place and rounding discussed here, but potentially to all cases where a crosslinguistic correlation can be found between the relative frequencies of feature values and the phonological strengths of these feature values.

5 Discussion

5.1 Summary

Computer simulations have shown that even if generation 1 has all faithfulness constraints top-ranked, we need assume no more than a fixed background (or transmission) noise to ensure that the learners of generation 2 will rank faithfulness by both average cue reliability (e.g. slightly higher for plosives than for nasals) and by frequency (e.g. slightly higher for labials than for coronals). Depending on how finely grained the faithfulness constraints themselves are contextualized, one will also find licensing by cue and/or positional faithfulness (e.g. faithfulness ranked higher for onsets or prevocalically than for codas or word-finally), leading to the well-known observed triple asymmetries (by place, manner, and position) in nasal place assimilation. These findings show that non-goal-oriented mechanisms (theoretically required by e.g. Ohala 1981 and Blevins 2004) can account for seemingly goal-oriented phenomena (which are observed facts).

5.2 Evolution over the generations

The shift between Generations 1 and 2 found in §3.6 does not lead to a language that is stable over the generations: when the outputs of Generations 2 and up are filtered with the same transmission noise as those of Generation 1, the auditory contrast between underlying labials and coronals will be washed out within a few generations.

To ensure stability over the generations, some simplifications may have to be undone: comprehension may have to be modelled as serial (McQueen & Cutler 1997), in which case the learning algorithm of parallel virtual production can be supplemented with lexicon-driven learning of prelexical perception (Boersma 1997, Escudero & Boersma 2003), which produces a prototype effect that ensures stability over the generations, at least if the articulatory constraint is split up appropriately (Boersma & Hamann 2007a). This connection to related learning algorithms has to be left for future research.

5.3 Crucial assumptions

Several hypotheses assumed in the present paper are crucial to make the explanations work.

One core hypothesis is that the grammar model contains at least three representations, as in Fig. 1. This is necessary to allow faithfulness to refer to discrete phonological elements such as [labial] and [coronal], while at the same time the transmission noise can influence continuous auditory detail in the phonetic form.

A similar core hypothesis is that of parallel production, as in Fig. 2. Without it, low-level articulatory constraints could not force the violation of faithfulness constraints, as in (4). If faithfulness constraints cannot be violated, they will not move up and down in learning tableaus such as (5), and they will not come to show any of the universal rankings.

Another core hypothesis is the multiple-level error-driven learning algorithm of (5). This type of learning algorithm was widely used in different contexts before (Tesar & Smolensky 2000, Apoussidou 2007), and is relatively uncontroversial. If in future this algorithm turns out to fail on different learning situations than humans do, the algorithm may have to be replaced. One indication for this is that although 91 percent of the virtual learners converged on the ranking in Fig. 3, the remaining 9 percent must therefore have failed. If we find 9 percent failures in human populations as well, this will be strong positive evidence that the learning algorithm is on the right track; if we do not find those failures in human populations, *something* in our modelling will have to be corrected.

A fourth hypothesis is the use of Optimality Theory as the general grammatical framework. We may try to trade it in for the framework of Harmonic Grammar (HG), in which constraints are weighted rather than ranked (Smolensky & Legendre 2006). However, when redoing the simulation of §4 with HG learners, they typically end up having their faithfulness constraints ranked by cue quality but not by frequency: IDENTPLACE will typically become ranked equally high for |p| as for |t|, and equally high for |m| as for |n|. This fact appears to provide evidence in favour of OT and against HG as models of human language processing, but as long as we do not know the precise cause of the difference in the present context, and as long as we do not know how this difference generalizes to other cases, it seems too early to tell for sure.

5.4 The bigger picture

The three representations in Fig. 1 are part of a larger model for bidirectional phonology and phonetics and their acquisition and evolution, which contains at least the following five representations: Morphemes – Underlying Form – Surface Form – Auditory Form – Articulatory Form. For the last three of these (surface, auditory, articulatory), Boersma & Hamann (2007a) found that lexicon-driven learning leads to an auditorily stable contrast between /s/ and /ʃ/ within a few generations, even if the starting language is unnatural. For the first three (morphemes, underlying, surface), Apoussidou (2007: ch.6) found that a bidirectional learning procedure fully analogous to the one discussed in the present paper leads to the automatic acquisition of hidden intermediate Underlying Forms that look as if linguists had devised them. The present paper, then, worked more or less with the middle three representations; it found that a bidirectional learning procedure leads to the automatic acquisition both of the hidden intermediate Surface Form as well as to phonologies (constraint rankings) with realistic biases. With all these ingredients, it looks as if Whole Language simulations of acquisition and evolution start to come within reach.

5.5 Phonologization, lexicalization, lexical selection

In the simulations of §3, Generation 1 started out by having no assimilation at all in their phonology or phonetics; at most we could say that they produced some auditory assimilation for their listeners. These listeners, however, heard this auditory assimilation and had to attribute it to some cause. Since nothing in the signal gave them any clue as to whether the cause was in the parents' phonology, the parents' phonetics, or the transmission medium, they distributed the cause automatically over the various candidates: they typically attributed full assimilation to the phonology and partial assimilation to the phonetics (§3.6).

We therefore have here a case of PHONOLOGIZATION: learners attribute the cause of an observed phenomenon to a higher level of processing than their adult teachers do.

This process can go a step further: if the input to production is the Morpheme (§5.4), and the selection of an Underlying Form given a Morpheme is performed in parallel with the phonology and the phonetics, learners may attribute observed phenomena to a yet higher level of processing, namely the selection of underlying forms, i.e. allomorph selection (Apoussidou 2007: ch.6). This is then a case of LEXICALIZATION. For place assimilation, this may have occurred for the Latin preposition *ad*, regarding its later restoration at the sentence level, at least in literary writing (e.g. *ad-* in *acclāmāre* versus *atque* and *ad caecum*).

And to go another step up in the representational hierarchy: the input to production could be a semantic representation, allowing that the selection of morphemes themselves is influenced by lower-level considerations. As conjectured by Boersma (1998: 382) and proved by Martin (2007: 53–69), such LEXICAL SELECTION processes in French restored the phoneme |b|, which had been missing in Proto-Indo-Eurorpean. Processes such as these may well have established the frequency differences between coronals and labials, or between rounded and unrounded vowels, i.e. the source of the "markedness" discussed in §4.3 (Boersma 2007c).

The more higher-level representations the grammar model contains, the more likely it is that the attribution process percolates up the representational hierarchy: from the phonetics into the phonology, into the morphology, and into the vocabulary. The mechanism described in the present paper might well be behind this whole chain of events.

6 Conclusion: what have we achieved and what not?

The present paper succeeded at explaining away some major aspects of markedness, by relating degrees of phonological strength directly to relative frequency. The paper also provided an underlying mechanism for the earlier proposals of positional faithfulness, licensing by cue, and markedness-by-faithfulness, suggesting that the earlier authors (Beckman 1998, Steriade 1995, Boersma 1998) were correct in ascribing their observed phenomena to faithfulness rankings but not in their proposed underlying causes (UG, auditory distance, listener-orientedness).

In sum, the paper seems to have reduced the potential relevance of no fewer than four previously proposed mechanisms. If these mechanisms turn out to be dispensible in general, this will suggest that the ultimate correct theory of language is less innatist and less functionalist than previously thought. After all, phenomena that emerge automatically within one language require no innate devices and no explicit goal orientation.

As for the techniques advanced in the present paper, room for improvement lies in the lack of evolutionary stability of the bare learning procedure (§5.2) and in the non-negligible percentage of learners who fail in some way (§3.7, §5.3). Future research can address these issues as well as investigate apparent counterexamples (§4.3) and applications to more complex instances of licensing by cue than could be handled here, such as the phenomena of cue trading (Steriade 1997). Finally, one would like to investigate whether the phonologization process observed in the simulations indeed has analogues throughout the grammar and the lexicon, as §5.5 suggests.

References

Apoussidou, Diana (2007). *The learnability of metrical phonology.* PhD thesis, University of Amsterdam.

Beckman, Jill N. (1998). *Positional faithfulness*. PhD thesis, University of Massachusetts, Amherst.

Blevins, Juliette (2004). *Evolutionary phonology*. Oxford University Press.

Boersma, Paul (1997). How we learn variation, optionality, and probability. *Proc. of the Institute of Phonetic Sciences* **21**: 43–58. University of Amsterdam.

Boersma, Paul (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. PhD thesis, University of Amsterdam.

Boersma, Paul (2003). Review of Tesar & Smolensky (2000): *Learnability in Optimality Theory*. *Phonology* **20**: 436–446.

Boersma, Paul (2006). Prototypicality judgments as inverted perception. In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds.) *Gradedness in grammar*. Oxford: Oxford University Press. 167–184.

Boersma, Paul (2007a). Some listener-oriented accounts of *h*-aspiré in French. *Lingua* **117**: 1989–2054.

Boersma, Paul (2007b). Cue constraints and their interactions in phonological perception and production. *Rutgers Optimality Archive* **944**.

Boersma, Paul (2007c). The evolution of phonotactic distributions in the lexicon. Talk presented at the Workshop on Variation, Gradience and Frequency in Phonology, Stanford, July 8, 2007.

Boersma, Paul, & Silke Hamann (2007a). The evolution of auditory contrast. *ROA* **909**.

Boersma, Paul, & Silke Hamann (2007b). Phonological perception in loanword adaptation. Talk presented at OCP 4, Rhodes. [full paper to be submitted for a book; deadline May 1, 2008]

Boersma, Paul, & Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* **32**: 45–86.

Boersma, Paul, & Joe Pater (2008). Convergence properties of a gradual learner in Harmonic Grammar. *Rutgers Optimality Archive* **xx**.

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

Escudero, Paola, & Paul Boersma (2003). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In Sudha Arunachalam, Elsi Kaiser & Alexander Williams (eds.) *Proceedings of the 25th Annual Penn Linguistics Colloquium. Penn Working Papers in Linguistics* **8.1**: 71–85. [Rutgers Optimality Archive **439**, 2001]

Escudero, Paola, & Paul Boersma (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* **26**: 551–585.

Flemming, Edward (1995). *Auditory representations in phonology*. Doctoral thesis, UCLA. [published 2002 by Routledge, New York & London]

Hume, Elizabeth, & Georgios Tserdanelis (2002). Labial unmarkedness in Sri Lankan Portuguese Creole. *Phonology* **19**: 441–58.

Hura, Susan L., Björn Lindblom & Randy L. Diehl (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech* **35**: 59–72.

Jun, Jongho (1995). Place assimilation as the result of conflicting perceptual and articulatory constraints. *WCCFL* **14**: 221–237.

Kirchner, Robert (1998). *Lenition in phonetically-based Optimality Theory.* Doctoral thesis, UCLA.

Maddieson, Ian (1984): *Patterns of sounds*. Cambridge: Cambridge University Press.

Malécot, André (1956). Acoustic cues for nasal consonants. *Language* **32**: 274–284.

Martin, Andrew T. (2007). *The evolving lexicon*. Doctoral thesis, UCLA.

McCarthy, John, & 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. 249–384.

McQueen, James M., & Anne Cutler (1997). Cognitive processes in speech perception. In William J. Hardcastle & John Laver (eds.) *The handbook of phonetic sciences*. Oxford: Blackwell. 566–585.

Menn, Lise (1983). Language acquisition, aphasia and phonotactic universals. In Fred Eckman, Edith Moravcsik & Jessica Wirth (eds.) *Markedness*. New York & London: Plenum Press. 241–256.

Mohanan, K.P. (1993). Fields of attraction in phonology. In John Goldsmith (ed.) *The last phonological rule: reflections on constraints and derivations*. University of Chicago Press. 61–116.

Mohr, B., & W.S.-Y. Wang (1968). Perceptual distance and the specification of phonological features. *Phonetica* **18**: 31–45.

Ohala, John J. (1981). The listener as a source of sound change. *CLS* **17**: 178–203.

Padgett, Jaye (1995). Partial class behavior and nasal place assimilation. *Proceedings of the South Western Optimality Theory Workshop*, Coyote Working Papers, University of Arizona, Tucson. 145–183.

Paradis, Carole, & Jean-François Prunet (1991). *The special status of coronals: internal and external evidence*. San Diego: Academic Press.

Pols, Louis C.W. (1983). Three-mode principal component analysis of confusion matrices, based on the identification of Dutch consonants, under various conditions of noise and reverberation. *Speech Communication* **2**: 275–293.

Prince, Alan, & Paul Smolensky (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Technical Report TR-**2**, Rutgers University Center for Cognitive Science. [published 2004 by Blackwell, Malden]

Smolensky, Paul, & Géraldine Legendre (eds., 2006). *The harmonic mind*. Cambridge, Mass.: MIT Press.

Steriade, Donca (1995). Positional neutralization. Two chapters of an unfinished manuscript, Department of Linguistics, UCLA.

Steriade, Donca (1997). Phonetics in phonology: the case of laryngeal neutralization. Unpublished manuscript, Department of Linguistics, UCLA. [version “June 1997”]

Steriade, Donca (2001). Directional asymmetries in place assimilation. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*. San Diego: Academic Press. 219–250.

Tesar, Bruce (1997). An iterative strategy for learning metrical stress in Optimality Theory. In Elizabeth Hughes, Mary Hughes, & Annabel Greenhill (eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development*, 615–626. Somerville, Mass.: Cascadilla.

Tesar, Bruce, & Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press.